SENTENCES UNDECIDABLE IN FORMALIZED ARITHMETIC

AN EXPOSITION OF THE THEORY OF KURT GÖDEL

ANDRZEJ MOSTOWSKI

Professor of Philosophy of Mathematics
University of Warsaw

1952

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM

PRINTED IN THE NETHERLANDS
DRUKKERIJ HOLLAND N.V., AMSTERDAM

PREFACE

In the present booklet an attempt is made to present as clearly and as rigorously as possible the famous theory of undecidable sentences created by Kurt Gödel in 1931.

The text of the booklet is based in part on an article published by the author in 1945 in the Polish periodical Kwartalnik Filozoficzny. In particular the Introduction is translated almost literally from that article. The rest of the text, however, has been completely changed and much new material has been added. The book therefore scarcely deserves the name of translation.

The theory of Gödel has been worked out by so many distinguished logicians in numerous books and special articles that the author feels compelled to excuse the publication of still one more book on the same subject. He therefore wishes to draw the attention of the reader to the theory of $\mathfrak{K}$-definability (Chapter V) which presents a simultaneous generalization of the theory of definability and that of the general recursivity of functions and relations. To the author's knowledge these theories have not been, until now, brought together.

In terms of the general theory of $\mathfrak{K}$-definable functions and relations it is possible to express clearly and conveniently the assumptions which are the common source of the various proofs of Gödel's incompleteness theorem formulated first by Gödel himself and then by Tarski and Rosser. It seems probable that the theory of $\mathfrak{K}$-definability will prove useful also in other logical researches.

To finish this preface the author whishes to express his gratitude to Professor E. W. Beth who invited him as early as in 1948 to contribute this booklet to the series "Studies in Logic" and to Dr Jan Kalicki and Mr R. G. Taylor for linguistic corrections of the text.

INTRODUCTION

A mathematical problem expressed in the form of a sentence S is called unsolved if we do not know whether the sentence S is true or false. Many of such unsolved problems are known in the contemporary mathematics. The most known belong to the theory of numbers. For instance the following three sentences express unsolved problems:

(i) *there are integers* x, y, z, n *such that* $n > 2$ *and* $x^n + y^n = z^n$ (Fermat's problem);

(ii) *there are infinitely many primes* p *such that* $p + 2$ *is prime* (problem of twins);

(iii) *there are infinitely many integers* n *such that* $2^{2^n} + 1$ *is prime.*

Mathematics is steadily progressing and solves incessantly problems which were left open by previous generations; simultaneously new problems sometimes of a considerable degree of difficulty arise and are being attacked by mathematicians.

It would be senseless to state of any of these problems that it is essentially unsolvable because we cannot predict what way will take the further development of mathematics, what new notions will be created by mathematicians and what new kinds of inference will be discovered.

We do not know whether there are problems which would remain unsolvable regardless of the new methods of correct proofs of whatever sort which may be found in the future. If such problems exist we can call them essentially unsolvable.

The question of existence of essentially unsolvable mathematical problems cannot be treated adequately before we possess an exact definition of essential unsolvability. This, however, is an enormous task, and we have at present no idea whether such a definition will ever be found.

We shall not deal with that problem any longer since it has no direct connection with our immediate purpose. We remark only that the difficulty, which we encounter when we try to formulate

the definition in question, has its source in the vagueness of the notion of a correct mathematical proof.

Let us assume for a moment that we succeeded to find an exact delimitation of this notion so that it is as clear for us as e.g. the notion of a prime or of an even integer.

The statement "a given problem S is (essentially) unsolvable" would then have a precise meaning independent of the present state of our knowledge in the domain of mathematics exactly as has the statement "each even integer is a sum of at most two primes". Anybody who trusts the laws of classical logic would also be entitled to maintain that each problem is either solvable or unsolvable, although the decision which case actually occurs for an arbitrarily proposed problem could be extremely difficult. The situation would be entirely similar to that which we have in arithmetic: we know (provided that we accept the ordinary logic) that each even integer n is either a sum of two primes or not; we do not know which of these both cases occurs for $n = 10^{10^{10^{10}}}$.

The development of the formal logic has led the scientists to constructions of formal systems which contain large parts of mathematics. Many logicians believed that some of these systems embrace the whole of mathematics, i.e. that each correct mathematical proof is formalizable within the systems.

The chief difference between the formal and the intuitive mathematics lies in the fact that the notions of a sentence and of a proof is exactly definable for the former but not for the latter. To describe a formal system we first enumerate signs from which expressions of the system can be built up, and we give exact rules how to do it. We also give further rules according to which certain finite sequences of expressions can be reckoned as proofs. And we define theorems (of the considered system) as expressions which can be the last terms of a proof. Hence the notions which were vague for the intuitive mathematics are quite precise for formalized mathematics and the notion of solvability of a problem (or of provability of a sentence) in a given formal system is perfectly clear and well defined.

Let us therefore consider a formal system (S) and a sentence S

written out by means which are at our disposal in (S). We may think of S as expressing a difficult mathematical problem, e.g. one of the problems (i)—(iii), whereas (S) can be e.g. a formalized system of arithmetic.

A priori three cases are possible for each S and (S): 1. S is a theorem of (S), 2. the negation of S is a theorem of (S), 3. neither 1 nor 2 occurs. In the cases 1 and 2 S is called decidable in (S), in the case 3 — undecidable in (S).

The system (S) is called complete if the case 3 never occurs, otherwise incomplete.

If all possible means of intuitive mathematics were expressible in (S), then a sentence undecidable in (S) would present an essentially unsolvable problem. Otherwise a sentence undecidable in (S) can very well be decidable in other more comprehensive systems and can be obviously false or obviously true in the intuitive mathematics. We see thus that until we succeed to build a formal system coinciding with the intuitive mathematics there is no immediate connection between the problem of completeness of any proposed formal system and the problem of existence of essentially unsolvable mathematical problems.

The problem of completeness of formalized systems is, however, important because it makes explicit the degree of difficulty of formalization of intuitive mathematics even if we restrict ourselves to that portion of mathematics which deals with integers. As we shall see in the sequel, no formalized system (S) can be complete if a certain well defined portion of arithmetic of integers is adequately expressible in (S). We shall show the incompleteness of such systems (S) constructing for each (S) an arithmetical sentence which is undecidable in (S). This undecidable sentence is — as we shall see — intuitively obvious and becomes provable when we strengthen the system (S) by addition of a number of intuitively obvious axioms and rules of inference.

It follows that no system (S) of the kind to which the method described below is applicable can coincide with the intuitive mathematics. It is important to note that the method is of such a degree of generality that it is applicable to practically all

systems (S) which deserve the name of formalized arithmetic [1].

The logicians of the first decades of the present century constructed a number of formal systems of logic in the hope to arrive finally at a system identical with the intuitive mathematics. If such a construction succeded, the notion of truth in mathematics would be definitely clarified, since it would become identical with the perfectly clear notion of provability in a formal system.

The theory due to Kurt Gödel, the main results of which we outlined above, shows how vain were these hopes. As a matter of fact, in spite of all efforts of the logicians we are still very far from an exact understanding in what consists the notion of truth in mathematics.

The purpose of the above remarks was to formulate the problem with which deals the theory of Gödel and to explain why this theory occupies the central place in mathematical logic. We shall now sketch the main lines of the argument which has led Gödel to his discoveries.

This argument is closely akin to the so called Richard's antinomy [2] and we shall therefore begin with a presentation of this antinomy. The reader should not think however, that Gödel's argument is itself an antinomy: Richard's argument contains a number of correct inferences and at the same time a number of very subtle, and not easily detectable errors which are responsible for the antinomy. Gödel corrected the errors of Richard and arranged the reasoning so that it can serve the purpose.

Richard's antinomy can be formulated as follows. We consider expressions of the English language which are definitions of properties of integers. The number of these expressions is of course denumerable and we can therefore arrange them in an infinite

[1] An exact formulation of conditions which a system (S) has to satisfy in order that the theory of incompleteness which we just described be applicable to it are given in Hilbert-Bernays [11], Vol. 2, p. 271. Numbers in brackets refer to the bibliography on p. 116.

[2] Richard [18].

sequence:

$$(1) \qquad W_1, W_2, \ldots, W_n, \ldots$$

We can agree e.g. that W_i precedes W_j if W_i contains less letters than W_j or — if both expressions have the same number of letters — if W_i precedes W_j in the lexicographical ordering of words.

For arbitrary integers n and p one of the following cases must occur: (i) n possesses the property expressed by W_p, (ii) n does not possess this property. We write in the case (i) $\vdash W_p(n)$ and in the case (ii) $\sim\vdash W_p(n)$.

Consider now the property of an integer n expressed by the formula

$$\sim\vdash W_n(n).$$

This property has been defined in the English language and must therefore coincide with one of the properties (1). Hence there is a q such that for each n the conditions $\vdash W_q(n)$ and $\sim\vdash W_n(n)$ are equivalent. Taking $n = q$ we obtain a contradiction: $\vdash W_q(q)$ is equivalent to $\sim\vdash W_q(q)$ [3].

We shall now try to carry over this antinomy from the unprecise everyday language to a formal system (S) which contains arithmetic of integers. Of course no reconstruction of the antinomy is possible if the system (S) is self-consistent. We shall see, indeed, that certain passages in the reasoning leading to the antinomy are not translatable to the language of (S). Modifying suitably these passages, so as to secure translatability, we obtain precisely the theorems of Gödel.

We consider all expressions of (S) which are definitions of properties of integers. In other words we consider matrices [4] of (S) with one free variable which runs over integers. The number of such matrices is evidently denumerable and we can arrange them in a sequence

$$(2) \qquad W_1, W_2, W_3, \ldots$$

[3] We do not need here to explain in what consists the error committed in this argument. See e.g. Fraenkel [6], p. 215, Carnap [1], p. 211–220.

[4] I.e. expressions which have the form of sentences but which differ from sentences by containing one or more free variables.

If we try to argue further as in the antinomy we encounter immediately a difficulty. We have to consider the following property of an integer n:

(3) *n does not possess the property expressed by W_n.*

Now, what does it mean that an integer possesses a property expressed by a matrix, or, as we also say, that an integer satisfies a matrix?

In order to discuss this question let us assume that to the signs of the system (S) belong signs **1**, **2**, **3**, ... which denote the consecutive integers. These signs will be called numerals and the n-th numeral will be denoted by **n**.

It is intuitively obvious that the integer n satisfies a matrix W if and only if the sentence $W(\mathbf{n})$ obtained from the matrix by a substitution of the n-th numeral for its unique variable is true. E.g. the number 2 satisfies the matrix

$$(y) \sim (y^2 = x)$$

since the sentence $(y) \sim (y^2 = 2)$ is true.

This remark, important though it is, is of little help because we do not know what a true sentence is. We remember however that for formalized systems we do possess a notion which is closely related to that of truth and which had to replace this notion according to the intentions of the first authors of formalized systems: namely the notion of formal provability.

This remark suggests that we can substitute for (3) the following property of n:

(4) *The sentence $W_n(\mathbf{n})$ is unprovable in* (S).

We do not maintain that (4) possesses exactly the same intuitive meaning as (3) [5]. At any rate (4) can be expressed with any desired

[5] As we have said above, Gödel's results show that the notion of provability does not correspond to that of intuitive mathematical truth. It is therefore to expect that (4) is not an adequate formulation of (3) and it can be shown that it is really so. This is, however, without immediate bearing on our problem.

degree of precision, whereas it is not clear how to express precisely the formula (3), although its intuitive meaning seems to be obvious.

The next step would consist in an identification of (4) with one of the properties expressed by matrices (2). This is again a difficulty. In (4) occur words like "sentence", "provability", "substitution" and so on. These words are names of notions which belong to the grammar of (S). On the contrary, (S) was a formalized system of mathematics and its constants denote mathematical notions such as integers and certain relations between them. It seems, therefore, that we have no right to maintain that the property (4) is identical (in extension) with one of the properties expressed by the matrices (2).

The following ingenious device discovered by Gödel shows, however, that this is not the case. This device is called the arithmetization of metamathematics, and can be properly compared with the introduction of numerical coordinates to the grammar [6].

Expressions of the system (S) are finite sequences of signs (variables and logical and mathematical constants). Hence if we let correspond to each sign an arbitrary integer we obtain automatically a correspondence between expressions and finite sequences of integers. Since there exist one – to – one correspondences between finite sequences of integers and integers themselves [7], we obtain a one – to – one correspondence between integers and expressions. In the same way we can establish a one – to – one correspondence between integers and finite sequences of expressions. Integers which correspond in this way to expressions or their sequences will be called their Gödel-numbers.

Because of the correspondence just described, to each class of expressions (or relation between expressions) there belongs a uniquely determined class of integers (or relation between integers). In many cases these classes or relations admit purely

[6] This remark is due in principle to Carnap [1], p. 12 (coordinate language).

[7] One such correspondence is given by the formula

$$n_1, \ldots, n_k \rightleftarrows 2^{n_1} . 3^{n_2} \ldots p_k^{n_k}$$

where p_k is the k-th prime.

arithmetical definitions which can be expressible in the formal system (S).

Let us analyze the property (4) from the point of view of arithmetization. The property (4) as it stands is defined with the use of grammatical terms. If we replace the grammatical notions by their arithmetical counterparts we obtain a new definition which is expressed in purely arithmetical terms, and identical (in extension) with the former. The situation is entirely similar to that which we have when we define an integer or a class of integers using geometrical notions such as points, collinearity etc. If we introduce arbitrary coordinates and replace the geometrical notions by their arithmetical counterparts expressible in arithmetic we obtain another definition of the same integer or of the same class. The reader will do well if he thinks this over in details expressing in two different ways the fact that there are at least $n = 4$ points not lying on the same plane: one definition is purely geometrical and another makes use of the coordinates.

Let $\varphi(n, p)$ be the Gödel-number of the sentence $W_n(\mathbf{p})$. This is a non-arithmetical definition of a function φ with two arguments but we shall see that it is equivalent to a purely arithmetical one which is expressible in (S), provided that (S) contains a sufficiently large portion of arithmetic. Let further T be the class of the Gödel-numbers of theorems of (S). Again this definition does not belong to arithmetic but can be shown to be equivalent with a purely arithmetical definition which is expressible in each sufficiently strong system (S).

The property (4) is equivalent to the following

$$(5) \qquad \varphi(n, n) \text{ non} \in T$$

and hence to an arithmetical definition expressible in (S).

It follows that there exists a matrix W_q of (S) which expresses in (S) the property (5). We can find explicitly this matrix when we write down in the language of the system (S) the arithmetical sentence (5).

We make now the final step and substitute the q-th numeral for the unique free variable of W_q. This gives us the sentence $W_q(\mathbf{q})$

which corresponds to the sentence constructed in the antinomy of Richard.

Let us try to understand the intuitive content of the sentence $W_q(\mathbf{q})$.It says that the number q has the property expressed by the matrix W_q. Now W_q was a formalization of (5), and hence $W_q(\mathbf{q})$ says the same as $\varphi(q, q)$ non $\in T$. In view of the equivalence of (4) and (5) the formula $\varphi(q, q)$ non $\in T$ means the same as: $W_q(\mathbf{q})$ is unprovable.

We have thus discovered the intuitive content of the sentence $W_q(\mathbf{q})$: It says that this very same sentence is unprovable in (S). This seems very paradoxical at first sight. But we can make our result more plausible if we observe that, owing to the arithmetization of metamathematics, there is for every sentence S of (S) an arithmetical sentence S', expressible in terms admitted in the system (S), which says that S is unprovable. There is nothing paradoxical in the fact that for a suitably chosen S, so to say by chance, the sentence S' turns out to be identical with S [8].

We ask now whether the sentence $W_q(\mathbf{q})$ can be decided in (S). If this sentence were provable, it would be intuitively false (because the content of $W_q(\mathbf{q})$ is that $W_q(\mathbf{q})$ is unprovable). But this is impossible because all theorems provable in (S) are intuitively true. If the negation of $W_q(\mathbf{q})$ were provable, then (under the assumption that (S) is a consistent system) $W_q(\mathbf{q})$ would be unprovable, and hence intuitively true. Hence the negation of an intuitively true sentence would be provable in (S) which is again impossible.

[8] The reader might ask why is it not possible to go further and to construct an S which satisfies the condition

S is true if and only if S is not true.

This would correspond exactly to the equivalence obtained in the Richard's paradox. It is not difficult to answer this question. In order to construct such an S it would be necessary to prove that the arithmetical counterpart of the (grammatical) statement *S is true* is expressible in (S). There are, however, no reasons to admit that this is really so. We possess as yet no formal definition of truth, and cannot expect that it will be possible to formulate this definition by the means available in (S). Cf. Chapter VI, section 1, corollary 2, p. 89.

The sentence $W_q(\mathbf{q})$ is thus undecidable in (S).

This is in outline the proof of Gödel's most important theorem.

Our proof is, of course, not yet complete and contains two essential gaps. First of all we assumed without proof that the function φ and the class T admit purely arithmetical definitions. This gap is relatively easy to fill although it requires a long list of rather laborious auxiliary notions. More important is the second gap: We used in our sketch of proof such an unclear notion as that of an intuitively true sentence and did not define what we really mean when we say that a definition is expressed in the system (S).

There are two different ways to overcome this essential difficulty. It is convenient to call the first semantical and the second syntactical.

If we choose *the semantical way, we try to give an exact definition* of what may be called the class of true sentences. We are of course not interested in that the definition be conform to this or other philosophical view on the nature of the notion of truth. We require only that the class (V) of what we call true sentences possesses the properties which are necessary for the proof outlined above.

Once the definition of the class (V) is given, it is an easy matter to define the other notions which are needed for the proof. So e.g. we say that a matrix W expresses in (S) the arithmetical property (5) of an integer n if the conditions

$$(6) \qquad W(\mathbf{n}) \textit{ is true} \quad \text{and} \quad \varphi(n, n) \text{ non} \in T$$

are equivalent for every n.

Not all properties of integers are expressible in (S) under this definition, but it can be shown that many important properties are so expressible, and that in particular the property (5) is among them.

Examining the sketch of the proof given above we see that the only properties of the class of true sentences required for the proof are the following:

(C_1) Every theorem of (S) is true,
(C_2) Negation of a theorem of (S) is never true,
(C_3) Conditions (6) are equivalent for each n.

Thus if we succeed to define a class (V) of sentences in such a way that the conditions (C_1), (C_2), and (C_3) will be satisfied, the proof of the Gödel's theorem will be accomplished [9].

The syntactical method tries to avoid the direct definition of truth and to arrange the proof so as to use exclusively the notion of provability. Using the syntactical way we shall have to say that a matrix W expresses a given property of integers $n_1, \ldots, n_k$ if it satisfies the two conditions

> *if* $n_1, \ldots, n_k$ *have the property, then* $W(\mathbf{n}_1, \ldots, \mathbf{n}_k)$ *is provable,*
> *if* $n_1, \ldots, n_k$ *do not have the property, then the negation of* $W(\mathbf{n}_1, \ldots, \mathbf{n}_k)$ *is provable* [10].

The class of properties which are expressible in (S) under this definition is considerably narrower than the class of properties expressible under the semantical definition and in particular it can be shown that the property (5) ceases now to be expressible in (S).

However, the following property of two integers

(7) *p is the Gödel number of a formalized proof of the sentence with the number $\varphi(n, n)$*

can easily the shown to be expressible in (S), also under the new definition. In other words, there exists a matrix W such that $W(\mathbf{p}, \mathbf{n})$ is provable in (S) if (7) holds, and the negation of $W(\mathbf{p}, \mathbf{n})$ is provable in (S) if (7) does not hold.

Evidently, there must exist a relationship between the property (5) of an integer n and the provability of the sentence

(8) $$\textit{for no } y \quad W(y, \mathbf{n})$$

which we shall note briefly as $(y) \sim W(y, \mathbf{n})$. Indeed, the formula

[9] The idea of the semantical proof of the incompleteness theorem is due to A. Tarski [22] and [23].

[10] $W(\mathbf{n}_1, \ldots, \mathbf{n}_k)$ denotes here evidently the sentence obtained from the matrix W by the substitution of the numerals $\mathbf{n}_1, \ldots, \mathbf{n}_k$ for the free variables occurring in the matrix.

(5) means that for no p does the formula (7) hold, and (7) is, as we have seen, intimately connected with the provability of the sentence $W(\mathbf{p}, \mathbf{n})$.

The relationship between the property (5) and the provability of the formula (8) can be expressed so:

If $\varphi(n, n) \in T$, *then* $\sim (y) \sim W(y, \mathbf{n})$ *is provable.*

Hence, if we take the matrix $(y) \sim W(y, x)$ as W_q and assume the consistency of (S) we can prove in exactly the same way as before that the sentence $W_q(\mathbf{q})$ whose Gödel number is $\varphi(q, q)$ is not provable in (S).

It follows by the properties of the matrix W that the sentences $\sim W(\mathbf{1}, \mathbf{q})$, $\sim W(\mathbf{2}, \mathbf{q}), \ldots, \sim W(\mathbf{m}, \mathbf{q}), \ldots$ are all provable in (S). If we assume that this implies the unprovability of $\sim (y) \sim W(y, \mathbf{q})$, we obtain the result that $\sim W_q(\mathbf{q})$ is not provable. The assumption just formulated is called the ω-consistency of (S). Hence we obtain the undecidability of $W_q(\mathbf{q})$ under the assumption of ω-consistency of (S)[11].

In Chapter VI (section 2) we shall learn still another variant of the proof, due to Rosser[12], which uses the syntactical method but assumes only the ordinary consistency of (S).

The different kinds of incompleteness proofs lead to different important corollaries. We obtain them when we investigate the problem of formalization of these proofs. It turns out that the semantical proof is not formalizable within (S) itself. As a corollary we obtain the important theorem that the notion of "truth" for the system (S) is not definable within (S). The syntactical proofs are on the contrary formalizable within (S) and studying carefully this fact we can recognize with Gödel that the consistency of (S) is not provable by means formalizable within (S). These and other corollaries to the theorem of Gödel will be sketched in the Appendix.

After these introductory explanations we pass now to the systematic exposition of the theory of Gödel. The plan of our study

[11] The proof here sketched is the original proof of Gödel [9].

[12] Rosser [19].

is the following: Chapter I brings the necessary facts of arithmetic of integers. In Chapter II we describe a formal system (S) for which the existence of undecidable sentences will be shown. The system will be presented simultaneously with its arithmetization. In Chapter III we prove many formal theorems in the system (S) and show in particular that this system is adequate for the arithmetic of integers. Chapter IV deals with the semantics of the system (S). Chapter V is devoted to the theory of representability of arithmetical functions and relations in the system (S). Finally, in Chapter VI and in the Appendix we use the material gained in the previous five Chapters to prove the Gödel's theorem together with various corollaries some of which we mentioned above.

I

AUXILIARY NOTIONS AND THEOREMS OF ARITHMETIC

1. Mapping of integers onto pairs and triples of integers. By *integers* we always understand the positive integers

$$1, 2, 3, 4, \ldots.$$

It is well-known that pairs (i, j) of integers can be arranged in a simple sequence, i.e., that there exists a one-to-one mapping of pairs (i, j) onto integers.

A simple function which effectuates this mapping is

$$J(i, j) = \tfrac{1}{2}(i + j - 1)(i + j - 2) + j.$$

Indeed, if we arrange the double array

$$\begin{array}{llll} (1, 1) & (1, 2) & (1, 3) & \ldots \\ (2, 1) & (2, 2) & (2, 3) & \ldots \\ (3, 1) & (3, 2) & (3, 3) & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{array}$$

into the sequence

(1) $\quad (1, 1)\ (2, 1)\ (1, 2)\ (3, 1)\ (2, 2)\ (1, 3) \ldots$

we obtain the pair (i, j) on the $J(i, j)$-th place in the sequence.

It follows easily that the function $J(i, j)$ satisfies the following two conditions:

For every k there are i, j such that $J(i, j) = k$,

If $J(i, j) = J(k, h)$, then $i = j$ and $k = h$.

These conditions are necessary and sufficient for the existence of functions converse to the function J, i.e. of two functions K_1 and K_2 such that $J(K_1(n), K_2(n)) = n$.

The pair $(K_1(n), K_2(n))$ occurs in the sequence (1) on the n-th place. It follows in particular that $K_1(1) = K_2(1) = 1$.

From the properties of the function J we easily obtain that the function

$$J_3(i, j, k) = J(i, J(j, k))$$

establishes a one-to-one correspondence between integers and triples of integers. Also the function J_3 possesses converse functions, i.e. there exist three functions L_1, L_2, L_3 such that

$$J_3(L_1(n),\ L_2(n),\ L_3(n)) = n.$$

The functions J and J_3 are increasing, i.e. if $a_1 \leqslant a_2$, $b_1 \leqslant b_2$, and $c_1 \leqslant c_2$, then $J(a_1, b_1) \leqslant J(a_2, b_2)$ and $J_3(a_1, b_1, c_1) \leqslant J_3(a_2, b_2, c_2)$.

Evidently, $L_1(1) = L_2(1) = L_3(1) = 1$.

We note finally the inequalities

$$J(a, b) < (a + b)^2 \text{ and } J_3(a, b, c) < (a + b + c)^4$$

which follow easily from the definition of the function J.

2. Mapping of integers onto finite sequences of integers. Since the set of all finite sequences of integers is denumerable, there exist mappings of integers onto finite sequences of integers. We shall define here a mapping which is perhaps not the simplest but which we shall find very useful in our further study.

We put

$$R(m, n) = 1 + m - n.\ [m/n]$$

where $[m/n]$ is the integral part of the ratio m/n. Thus $R(m, n) - 1$ is the remainder after division of m by n.

If $R(m, n) = 2$, then $R(mp, n) = R(p, n)$. Indeed $R(m, n) = 2$ implies that $m = kn + 1$ whence $mp = knp + p$ which shows that when dividing mp by n we get the same remainder as when dividing p by n.

We shall now prove the following lemma [1]

Lemma 1. *If integers $d_1, \ldots, d_m$ are pairwise prime and are all less than d and if $d_i > a_i$ for $i = 1, 2, \ldots, m$, then there is an integer $e < d^{d+4}$ such that $R(e, d_i) = a_i$ for $i = 1, 2, \ldots, m$.*

[1] Apart from the evaluation of e, this lemma is identical with the so called Chinese remainder theorem. Cf. Dickson [5], p. 11.

Proof. Put $p = d_1 d_2 \ldots d_m$, $p_i = p/d_i$ $(i = 1, \ldots, m)$. Hence p_i and d_i are relatively prime and it follows by the elements of the theory of numbers that there exists an integer x_i such that

$$\begin{aligned} & d_i < x_i < 2d_i \\ (1) \qquad & x_i p_i \equiv 1 \pmod{d_i} \text{ i.e., } R(x_i p_i, d_i) = 2. \end{aligned}$$

Put

$$e = x_1 p_1 (a_1 - 1) + \ldots + x_m p_m (a_m - 1).$$

Since p_j is divisible by d_i for $i \neq j$, it follows that the remainder after division of e by d_i is the same as the remainder resulting from division of $x_i p_i (a_i - 1)$ by d_i. Hence $R(e, d_i) = R(x_i p_i (a_i - 1), d_i)$ and we obtain using (1)

$$R(e, d_i) = R(a_i - 1, d_i) = 1 + (a_i - 1) + [(a_i - 1)/d_i] = a_i$$

because the integral part of $(a_i - 1)/d_i$ is 0.

The evaluation of e is obtained as follows:

Since $p_i < d^m$, $a_i < d$, $m < d$, and $x_i < 2d$, we obtain

$$e < m . 2d . d^d . d < d . d^2 . d^d . d = d^{d+4}.$$

Lemma 2[2]. *If* $a_1, \ldots, a_m$ *are arbitrary integers and* a *is such that* $m < a$ *and* $a_i < a$ *for* $i = 1, 2, \ldots, m$, *then there are integers* e *and* f *such that*

$$f < a^a, \ e < a^{a^{a+1}+4}$$

and

$$R(e, 1 + if) = a_i \text{ for } i = 1, 2, \ldots, m.$$

Proof. Put $f = a!$ We have then evidently

$$1 + (a - 1) f < a^a \text{ and } f < a^a.$$

We shall show that the integers $1 + if$ and $1 + jf$ are relatively prime for $i < j < a$. Indeed, suppose there is a prime p which divides $1 + if$ and $1 + jf$. Hence p divides the difference $(1 + jf) + - (1 + if) = (j - i)f$, i.e., it divides either $j - i$ or f. Since however

[2] Cf. Gödel [9], p. 192.

$j - i$ is less than a, it divides $a!$ and therefore p divides f in both cases. We obtain thus a contradiction because no prime divisor of f can be a divisor of $1 + if$.

We apply now lemma 1 taking $d_i = 1 + if$. Since $d_i \leqslant 1 + (a-1) f < a^a$ for $i = 1, 2, \ldots, m$, we obtain an integer e such that $R(e, 1 + if) = a_i$ for $i = 1, 2, \ldots, m$ and $e < (a^a)^{a^a+4} = a^{a^{a+1}+4}$.

The proof of the lemma 2 is thus complete.

We put now for every integer g

$$L(g) = L_3(g),\ \bar{g}_i = R(L_1(g),\ 1 + iL_2(g)),\ \bar{g}_* = \bar{g}_{L(g)}.$$

We let correspond to every g the finite sequence

$$\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_{L(g)} = \bar{g}_*$$

with $L(g)$ terms. We shall say that the integer g represents this sequence.

Theorem 3. *For every sequence $a_1, a_2, \ldots, a_m$ there is an integer g representing this sequence i.e., satisfying the conditions*

$$L(g) = m,\ \bar{g}_i = a_i \text{ for } i = 1, 2, \ldots, m.$$

If $m < a$ and $a_i < a$ for $i = 1, 2, \ldots, m$, then g can be found among numbers less than

$$T(a) = J_3(a^{a^{a+1}+4}, a^a, a).$$

Proof. We determine integers e and f as in lemma 2 and put $g = J_3(e, f, m)$. Since J_3 is an increasing function and $m < a$, we obtain

$$g < J_3(a^{a^{a+1}+4}, a^a, a) = T(a).$$

It follows further from the definition of g that

$$\bar{g}_i = R(L_1(g), 1 + iL_2(g)) = R(e, 1 + if) = a_i$$

for $i = 1, 2, \ldots, m$, which proves the theorem.

3. **Logical symbols.** It will be convenient to abbreviate some phrases and write them with the help of symbols. We shall use the

symbols

$$.,\ \mathbf{v},\ \supset,\ \equiv,\ \sim,\ (\exists n),\ (n)$$

instead of the words *and*, *or*, *if . . . then*, *if and only if*, *not*, *there is an integer n such that*, *for every integer*.

If F is a function from integers to integers, we write

(1) $$(\exists n)_{F(a,b,\ldots,k)},\ (n)_{F(a,b,\ldots,k)}$$

instead of *there is an integer n less than* $F(a, b, \ldots, k)$ and *for every integer n less than* $F(a, b, \ldots, k)$.

The symbols (1) are called the limited quantifiers. They are particularly useful as abbreviations of alternations and conjunctions containing many terms. E.g. we can write $(\exists n)_k[\ldots n \ldots]$ instead of

$$[\ldots 1 \ldots]\ \mathbf{v}\ [\ldots 2 \ldots]\ \mathbf{v}\ \ldots\ \mathbf{v}\ [\ldots k-1 \ldots]$$

and similarly for conjunctions with k terms.

4. Relations and functions. The general notion of relations and functions will be assumed as known. We shall consider only relations between integers and functions whose values and arguments range over the set of integers. Relations will be denoted by German letters and functions by Roman capitals.

Relations are divided into singulary, binary, ternary, and so on. For instance the less-than-relation is binary; the relation defined by the formula $m = n + p$ is ternary. Singulary relations can be identified with sets of integers.

If $\mathfrak{R}$ is a k-ary relation, we write $\mathfrak{R}(n_1, \ldots, n_k)$ in order to indicate that $\mathfrak{R}$ holds between the integers $n_1, \ldots, n_k$. If $k = 1$, we write usually $n \in \mathfrak{R}$ instead of $\mathfrak{R}(n)$.

We shall use the Greek letter λ as the relational abstract. In other words, if we have any condition on integers

$$\ldots x \ldots y \ldots z \ldots,$$

then we denote by $\lambda xyz\ldots\ [\ldots x \ldots y \ldots z \ldots]$ the relation which holds between the integers $x, y, z, \ldots$ if and only if they satisfy the given condition.

The set whose unique elements are $a_1, a_2, \ldots, a_m$ will be denoted by $\{a_1, a_2, \ldots, a_m\}$.

Functions are classified according to the number of their arguments. $F(n_1, \ldots, n_k)$ denotes as usual the value of the function F for the arguments $n_1, \ldots, n_k$.

We shall now describe a few operations which lead from given relations or functions to other relations or functions. We assume that $\mathfrak{R}$ and $\mathfrak{S}$ are an m-ary and an n-ary relations and that F is a function with k arguments.

a. Boolean operations. The *complement* of $\mathfrak{R}$ is defined as the m-ary relation

$$-\mathfrak{R} = \lambda a_1 \ldots a_m[\sim \mathfrak{R}(a_1, \ldots, a_m)]$$

and the (Boolean) *sum* of $\mathfrak{R}$ and $\mathfrak{S}$ as the $m+n$-ary relation

$$\mathfrak{R} \vee \mathfrak{S} = \lambda a_1 \ldots a_m b_1 \ldots b_n [\mathfrak{R}(a_1, \ldots, a_m) \vee \mathfrak{S}(b_1, \ldots, b_n)].$$

Other Boolean operations on relations are definable in terms of sums and complements, e.g. the *product* (intersection) of $\mathfrak{R}$ and $\mathfrak{S}$ and the *difference* of $\mathfrak{R}$ and $\mathfrak{S}$

$$\mathfrak{R} \wedge \mathfrak{S} = -[-\mathfrak{R} \vee -\mathfrak{S}], \quad \mathfrak{R} - \mathfrak{S} = \mathfrak{R} \wedge -\mathfrak{S},$$

and several others. Of special importance is the relation $-\mathfrak{R} \vee \mathfrak{S}$ which holds between the integers $a_1, \ldots, a_m, b_1, \ldots, b_n$ if and only if $\mathfrak{R}(a_1, \ldots, a_m) \supset \mathfrak{S}(b_1, \ldots, b_n)$.

If $m = n = 1$ we shall sometimes denote by $\mathfrak{R} \vee \mathfrak{S}$ and $\mathfrak{R} \wedge \mathfrak{S}$ the set-theoretical union and intersection of $\mathfrak{R}$ and $\mathfrak{S}$ (i.e. the arguments of $\mathfrak{R}$ and $\mathfrak{S}$ will be identified).

b. The operation $\min_j$. If $1 < j < m$, we denote by $\min_j \mathfrak{R}$ the function F with $m-1$ arguments defined as follows: If there is an integer a_j such that $\mathfrak{R}(a_1, \ldots, a_m)$, then

$$F(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m)$$

is the smallest such integer a_j. Otherwise the value of F is 1.

c. Identification of arguments. This operation leads from a relation $\mathfrak{R}$ or a function F to relations or functions with a smaller number of arguments. For instance, the identification of the i-th

and the j-th arguments leads from $\mathfrak{R}$ to the relation

$$\lambda a_1 \ldots a_{j-1}\, a_{j+1} \ldots a_m\, [\mathfrak{R}(a_1 \ldots a_{j-1}\, a_i\, a_{j+1} \ldots a_m)].$$

d. Substitution. The following relation is said to arise from $\mathfrak{R}$ by a substitution of F for the i-th argument of $\mathfrak{R}$:

$$\lambda a_1 \ldots a_{i-1}\; a_{i+1} \ldots a_m b_1 \ldots b_k\, [\mathfrak{R}(a_1, \ldots, a_{i-1}, F(b_1, \ldots, b_k), a_{i+1}, \ldots, a_m)].$$

The operation of substitution for functions is defined similarly.

In terms of these operations we define still the following ones:

$$\mathfrak{R}^{(j)} = \lambda a_1 \ldots a_{j-1} a_{j+1} \ldots a_m [\mathfrak{R}(a_1, \ldots, a_{j-1}, \min_j \mathfrak{R}(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m), a_{j+1}, \ldots, a_m)],$$

$$\overline{\mathfrak{R}}^{(j)} = -\{[-\mathfrak{R}]^{(j)}\}.$$

It can be shown easily that

$$\mathfrak{R}^{(j)}(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m) \equiv (\exists a_j)\mathfrak{R}(a_1, \ldots, a_m),$$

$$\overline{\mathfrak{R}}^{(j)}(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m) \equiv (a_j)\mathfrak{R}(a_1, \ldots, a_m).$$

The operations leading from $\mathfrak{R}$ to $\mathfrak{R}^{(j)}$ and $\overline{\mathfrak{R}}^{(j)}$ correspond thus to the existential and the general quantifiers. We can also consider similar operations which correspond to the limited quantifiers:

$$\mathfrak{R}^{(j)}_F = \lambda a_1 \ldots a_{j-1} a_{j+1} \ldots a_m b_1 \ldots b_k\, [(\exists a_j)_{F(b_1, \ldots, b_k)} \mathfrak{R}(a_1, \ldots, a_m)],$$

$$\overline{\mathfrak{R}}^{(j)}_F = \lambda a_1 \ldots a_{j-1} a_{j+1} \ldots a_m b_1 \ldots b_k\, [(a_j)_{F(b_1, \ldots, b_k)} \mathfrak{R}(a_1, \ldots, a_m)].$$

Also these operations are definable in terms of the operations a — d and the constant relations $=$ and $<$. Indeed, if we put

$$\mathfrak{S} = \lambda a_1 \ldots a_m b_1 \ldots b_k\, [\mathfrak{R}(a_1, \ldots, a_m) \vee a_j = F(b_1, \ldots, b_k)],$$

$$G = \min_j \mathfrak{S},$$

we easily see that

$$\mathfrak{R}^{(j)}_F(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m, b_1, \ldots, b_k) \equiv$$
$$[\mathfrak{R}(a_1, \ldots, a_{j-1}, G(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m, b_1, \ldots, b_k), a_{j+1}, \ldots, a_m) \cdot$$
$$G(a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m, b_1, \ldots, b_k) < F(b_1, \ldots, b_k)],$$

and $\overline{\mathfrak{R}}^{(j)}_F = -\{[-\mathfrak{R}]^{(j)}_F\}$.

Note that the operation $\min_j$ has been performed here on a relation $\mathfrak{S}$ which satisfies the condition

$$(a_1)(a_2)\ldots(a_{j-1})(a_{j+1})\ldots(a_m)(b_1)\ldots(b_k)(\exists a_j) \\ \mathfrak{S}(a_1, \ldots, a_m, b_1, \ldots, b_k).$$

5. Inductive definitions. A typical inductive definition of a function with, say, 2 arguments has the form [3]

$$(1_1) \qquad F(1, y) = G(y),$$

$$(1_2) \qquad F(x+1, y) = H(x, y, F(x, y))$$

where G and H are known functions.

We shall show here how to transform this inductive definition into an ordinary (explicit) one [4].

It is easy to show that $u = F(n, y)$ if and only if there exists a sequence

$$(2) \qquad u_1, u_2, \ldots, u_n = u$$

such that $u_1 = G(y)$ and $u_{i+1} = H(i, y, u_i)$ for $i = 1, 2, \ldots, n-1$. Now let g be an integer which represents the sequence (2). Every such integer satisfies the conditions

$$(3) \quad L(g) = n,\ \bar{g}_1 = G(y),\ (i)_n[\bar{g}_{i+1} = H(i, y, \bar{g}_i)],\ F(n, y) = \bar{g}_n.$$

Hence, if we introduce the abbreviation

$$\mathfrak{R} = \lambda gny[(L(g) = n)\ .\ (\bar{g}_1 = G(y))\ .\ (i)_n[\bar{g}_{i+1} = H(i, y, \bar{g}_i)]],$$

we shall have

$$(4) \qquad F(n, y) = \overline{\min_1 \mathfrak{R}(n, y)}_*$$

because $\min_1 \mathfrak{R}(n, y)$ is the least g such that $\mathfrak{R}(g, n, y)$, i.e. such that the conditions (3) are satisfied.

[3] This is the schema of the so called primitive recursion. Other types of inductive definitions can be found e.g. in Hilbert–Bernays [11] vol. 1, pp. 325–343.

[4] The origin of this transformation goes back to the classical investigations of Frege [7] concerning the logical definition of integers and of Dedekind [4] concerning chains. Cf. further Gödel [9], p. 191, Satz VII.

Note that $\mathfrak{R}$ satisfies the condition

$$(n)(y)(\exists g)\ \mathfrak{R}(g, n, y).$$

Definition (4) is an ordinary (explicit) definition of a function. It is equivalent to the definition (1) in the sense that the function defined by equation (4) satisfies the equations (1_1) and (1_2).

Indeed, assume first that $n = 1$. If $\mathfrak{R}(g, 1, y)$, then $L(g) = 1$ and $\bar{g}_1 = G(y)$. Since there are integers g satisfying these equations, $\min_1 \mathfrak{R}(1, y)$ is one of them. Hence $\overline{\min_1 \mathfrak{R}(1, y)}_* = G(y)$ which proves that the function (4) satisfies the equation (1_1).

The proof of (1_2) is a little more involved and requires two lemmas.

Lemma 1. *For every n there are integers g such that $\mathfrak{R}(g, n, y)$.*

Proof. As we have shown above, the lemma holds for $n = 1$. We show that if it holds for an integer n, it does so for the integer $n + 1$. Indeed, let g be such that $\mathfrak{R}(g, n, y)$ and let g' represent the sequence

$$\bar{g}_1, \bar{g}_2, \ldots, \bar{g}_n, H(n, y, \bar{g}_n). \tag{5}$$

We have then $L(g') = n + 1$ because the sequence (5) has $n + 1$ terms. If $i + 1 \leqslant n$, then $\bar{g}'_{i+1} = \bar{g}_{i+1}$, $\bar{g}'_i = \bar{g}_i$, and hence

$$\bar{g}'_{i+1} = H(i, y, \bar{g}'_i)$$

because of the properties of g. The same equation holds for $i = n$, since $\bar{g}'_n = \bar{g}_n$ and $\bar{g}'_{n+1}$ is by definition equal to $H(n, y, \bar{g}_n)$.

Thus we have proved that $(i)_{n+1}\,[\bar{g}'_{i+1} = H(i, y, \bar{g}'_i)]$ and since $\bar{g}'_1 = \bar{g}_1 = G(y)$, we obtain $\mathfrak{R}(g', n + 1, y)$, q.e.d.

Lemma 2. *If $\mathfrak{R}(g, n, y)$ and $\mathfrak{R}(g', m, y)$ where $m > n$, then $\bar{g}_i = \bar{g}'_i$ for $i = 1, 2, \ldots, n$.*

Proof. For $i = 1$ the equation is evident since $\bar{g}_1 = \bar{g}'_1 = G(y)$. If $i < n$ and $\bar{g}_i = \bar{g}'_i$, then

$$\bar{g}_{i+1} = H(i, y, \bar{g}_i),$$
$$\bar{g}'_{i+1} = H(i, y, \bar{g}'_i),$$

and hence $\bar{g}_{i+1} = \bar{g}'_{i+1}$, q.e.d.

We can now prove that the function (4) satisfies the formula (1_2). Put $g = \min_1 \Re(n, y)$, $g' = \min_1 \Re(n + 1, y)$. Hence we obtain $\Re(g, n, y)$ and $\Re(g', n + 1, y)$ according to the lemma 1, and therefore $\bar{g}_n = \bar{g}'_n$ by lemma 2. Since

$$\bar{g}'_{n+1} = H(n, y, \bar{g}'_n),$$

it follows that $\bar{g}'_{n+1} = H(n, y, \bar{g}_n)$. But $\bar{g}_n = F(n, y)$, $\bar{g}'_{n+1} = F(n + 1, y)$ by the definition (4). Hence we obtain

$$F(n + 1, y) = H(n, y, F(n, y)),$$

as was to be shown.

Our discussion shows that the explicit definition (4) can replace the inductive definition (1) since the function defined in (4) possesses the properties required in (1). Other examples of elimination of inductive definitions will be given in Chapter II.

II

THE SYSTEM (S) AND ITS SYNTAX

In the present Chapter we give a detailed description of a formal system of arithmetic.

We shall be interested exclusively in the formal properties of expressions of the system and not in their intuitive content, i.e., we shall study operations on and relations between expressions treated as strings of signs deprived of every meaning whatsoever. Since however it is much easier to pursue this formal study with a definite image before the eyes, we shall explain briefly what meaning may be attached to the expressions of the system. These explanations are, of course, entirely informal and we shall never refer to them in the later parts of our exposition.

1. Preliminary description of the system. The signs from which the expressions of (S) are built up are the following:

free variables $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n, \ldots$,
bound variables $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \ldots$,
the arithmetic constants **1**, **sum**, **mult**,
the sign of equality **equ**,
the propositional connective **imp**,
the least-number operator **min**.

The sign **1** is to be interpreted as a name of the integer 1 and variables as names of unspecified integers. The signs **sum** and **mult** denote the usual addition and multiplication of integers. If we combine the variables and the sign **1** with the arithmetical signs **sum** and **mult** (and in cases of need with parentheses), we get expressions which we shall call the simplest *functional forms*. They correspond approximatively to what a mathematician would call polynomials with positive integral coefficients.

Putting the sign of equality between functional forms, we get the simplest *matrix-forms*. Other matrix-forms can be obtained

from the simplest ones by putting the sign of implication (to be read: if . . ., then . . .) between two matrix-forms.

The operator **min** serves to construct the expressions which denote least integers satisfying given properties. Thus, e.g., if E is an equation both sides of which are polynomials in the variable $\mathbf{v}_i$, then $(\mathbf{min}\ \mathbf{v}_i)\ E$ denotes the smallest integral root of this equation or 1 if the equation has no integral roots. In general, E contains also variables different from $\mathbf{v}_i$ and the smallest root of E is a function of these variables. Hence $(\mathbf{min}\ \mathbf{v}_i)\ E$ denotes an arithmetical function in the same sense as does the expression $\mathbf{a}_1$ **sum** $\mathbf{a}_2$.

Starting with the functional forms built up with the help of the sign **min**, and applying to them the arithmetical operations **sum** and **mult** we get a new kind of functional forms. These forms combined with the sign of equality and the propositional connective give rise to a new kind of matrix-forms. The **min**-operator applied to matrices of the new kind gives new functional forms and so on.

These considerations suffice to characterize the portion of arithmetic which can be expressed in the system (S). We shall describe below matrices which are admitted as axioms of the system (S), and shall enumerate operations which allow us to obtain theorems from the axioms. It will then be seen that many intuitive arithmetical proofs can be carried over to the system (S). This system contains thus a considerable part of arithmetic.

We remark still that the distinction which we made above between free and bound variables is not essential. It is convenient, however, to use the **min**-operator exclusively in connection with one set of variables, whereas the operation of substitution, which we shall describe in section **3**, is to be performed exclusively on the variables of the second set.

We proceed now to a formal description of the system.

2. Expressions and integers. In order to be able to speak about the signs enumerated in **1**, we must possess names for them[1].

[1] Speaking generally one must use names of things when one wishes to make statements about these things. Only exceptionally can a thing be its own name.

Furthermore, we must assume without proof certain fundamental properties of signs and their combinations (e.g., that no combination of two signs is itself a sign, that a variable is never an arithmetical sign nor a connective etc). In this way we arrive at a theory of expressions which is in principle not different from any other mathematical theory. This theory of expressions of a formal system is called after Hilbert the meta-system [2].

Developing the meta-system one has often to use notions which belong to quite advanced parts of mathematics. Thus, the meta-system is composed of two parts one of which contains exclusively general notions of logic and mathematics, and the other, specific notions of the meta-system.

Since we wish to avoid as far as possible any duplication of proofs, we shall assume a rather strong formulation of the axioms dealing with the second part of the meta-system. We shall namely assume that expressions can be mapped in a one-to-one manner on a set of integers [3]. This axiom will enable us to dispense with expressions altogether: Instead of expressions we shall deal uniquely with the integers which correspond to them [4]. Our exposition will therefore belong entirely to arithmetic. However, in view of the axiom of the meta-system which we just mentioned, every arithmetical theorem can be interpreted as a theorem concerning expressions.

It should still be noted that the arithmetic which we shall use is the ordinary intuitive arithmetic known to everybody from the school. Of course, that part of arithmetic which we need for our purposes can be axiomatized, and it is an important question whether we obtain then a system equivalent to (S) or stronger than (S). We shall discuss this problem in the Appendix. For the

[2] Axioms for a deductive theory of expressions were first proposed by Tarski [22], p. 289.

[3] This follows also from the axioms of the deductive meta-system. Cf. Tarski [22], pp. 301–302.

[4] This method of discussing properties of expressions is due to Gödel [9]. In principle, it would be also possible to take an opposite course and consider arithmetic as a part of a suitable meta-system. Cf. Quine [17].

present however, we use arithmetic with the same freedom as in any mathematical textbook and accept in particular proofs and definitions by induction[5]. Lower case Latin letters will be used as variables in the intuitive arithmetic. Their range is the set of positive integers.

Before we formulate the fundamental axiom of the meta-system, we define certain auxiliary functions and constants:[6]

$$\begin{aligned}
&\mathrm{v}_h = 4h, \quad \mathrm{a}_h = 4h+1 \quad h = 1, 2, 3, \ldots \\
&\mathrm{x} = \mathrm{v}_1, \quad \mathrm{y} = \mathrm{v}_2, \quad \mathrm{z} = \mathrm{v}_3, \quad \mathrm{t} = \mathrm{v}_4, \\
&\mathrm{a} = \mathrm{a}_1, \quad \mathrm{b} = \mathrm{a}_2, \quad \mathrm{c} = \mathrm{a}_3, \quad \mathrm{d} = \mathrm{a}_4, \\
&a \mathbin{\boldsymbol{+}} b = 4J_3(1, a, b) + 2, \\
&a \times b = 4J_3(2, a, b) + 2, \\
&a \approx b = 4J_3(3, a, b) + 3, \\
&a \rightarrow b = 4J_3(4, a, b) + 3, \\
&(\mu \mathrm{v}_h)a = 4J_3(5, h, a) + 2.
\end{aligned}$$

We shall use the words free variable, printed in spaced characters, instead of "an integer of the form a_h", and the words bound variable instead of "integer of the form v_h". The integers $a \mathbin{\boldsymbol{+}} b$ and $a \times b$ will be called the sum and the product of a and b. The integer $a \approx b$ will be called the equation between a and b, and the integer $a \rightarrow b$ the implication with the antecedent a and the consequent b. Finally, the integer $(\mu \mathrm{v}_h)a$ will be said to arise from a by an application of the μ-operator.

The set of bound variables will be denoted by $\mathfrak{Vb}$ and the set of free variables by $\mathfrak{Vf}$:

$$a \in \mathfrak{Vb} \equiv R(a, 4) = 1, \quad a \in \mathfrak{Vf} \equiv R(a, 4) = 2.$$

We define now by a simultaneous induction two sets of integers $\mathfrak{Ff}$ and $\mathfrak{Mf}$ (called the set of functional forms and the set of matrix-forms respectively) and three functions Oc, Ind, and S.

[5] Cf. Landau [14].

[6] The actual form of these functions is not essential. All we need is that two different functions never take on the same value and that the value of the function determines the values of the arguments.

The function Oc will be capable of but two values 1 and 2. The formula $Oc(a, b) = 2$ will be read: a occurs in b, and the formula $Oc(a, b) = 1$: a does not occur in b. $Ind(i, b)$ will be called the index of the bound variable v_i in b, and $S(i, p, b)$ will be called the result of substitution of p for the variable a_i throughout b.

(1) $1 \in \mathfrak{F}\mathfrak{f}$; $Oc(a, 1) = 1$; $Ind(i, 1) = 1$ for every i; $S(i, p, 1) = 1$ for every i.

(2) $\mathrm{a}_h \in \mathfrak{F}\mathfrak{f}$; $Oc(a, \mathrm{a}_h) = 2$ or 1 according as $a = \mathrm{a}_h$ or $a \neq \mathrm{a}_h$; $Ind(i, \mathrm{a}_h) = 1$ for every i; $S(i, p, \mathrm{a}_h) = \mathrm{a}_h$ if $i \neq h$ and $S(h, p, \mathrm{a}_h) = p$.

(3) $\mathrm{v}_h \in \mathfrak{F}\mathfrak{f}$; $Oc(a, \mathrm{v}_h) = 2$ or 1 according as $a = \mathrm{v}_h$ or $a \neq \mathrm{v}_h$; $Ind(i, \mathrm{v}_h) = 1$ for $i \neq h$ and $Ind(h, \mathrm{v}_h) = 2$; $S(i, p, \mathrm{v}_h) = \mathrm{v}_h$ for every i.

(4) If b and c are in $\mathfrak{F}\mathfrak{f}$, then so are $b + c$ and $b \times c$ and $b \approx c$ is in $\mathfrak{M}\mathfrak{f}$; $Oc(a, b + c) = \max\,(Oc(a, b), Oc(a, c))$ and similarly for $b \times c$ and $b \approx c$; $Ind(i, b + c) = \max\,(Ind(i, b), Ind(i, c))$ for every i and similarly for $b \times c$ and $b \approx c$; $S(i, p, b + c) = S(i, p, b) + S(i, p, c)$ and similarly for $b \times c$ and $b \approx c$.

(5) If b and c are in $\mathfrak{M}\mathfrak{f}$, then so is $b \rightarrow c$; $Oc(a, b \rightarrow c) = \max\,(Oc(a, b), \; Oc(a, c))$; $Ind(i, b \rightarrow c) = \max\,(Ind(i, b), Ind(i, c))$; $S(i, p, b \rightarrow c) = S(i, p, b) \rightarrow S(i, p, c)$.

(6) If b is in $\mathfrak{M}\mathfrak{f}$, then $(\mu \mathrm{v}_j)b$ is in $\mathfrak{F}\mathfrak{f}$; $Oc(a, (\mu \mathrm{v}_j)b) = 2$ if and only if $a = \mathrm{v}_j$ or $Oc(a, b) = 2$, otherwise the value of $Oc(a, (\mu \mathrm{v}_j)b)$ is 1; $Ind(i, (\mu \mathrm{v}_j)b) = Ind(i, b)$ if $i \neq j$ and $Ind(j, (\mu \mathrm{v}_j)b) = 1$; $S(i, p, (\mu \mathrm{v}_j)b) = (\mu \mathrm{v}_j)S(i, p, b)$.

In order to explain how this definition works, we shall show that the integer

$$\Phi = (\mu \mathrm{z})[(\mathrm{z} + \mathrm{a}) \approx (\mathrm{a} + 1)] \approx \mathrm{x}$$

is in $\mathfrak{M}\mathfrak{f}$. Indeed $\mathrm{z} + \mathrm{a}$ and $\mathrm{a} + 1$ are in $\mathfrak{F}\mathfrak{f}$ by (1), (2), (3), and (4). Hence by (4) $\mathrm{z} + \mathrm{a} \approx \mathrm{a} + 1$ is in $\mathfrak{M}\mathfrak{f}$ and therefore by (6) $(\mu \mathrm{z})[(\mathrm{z} + \mathrm{a}) \approx \mathrm{a} + 1]$ is in $\mathfrak{F}\mathfrak{f}$. Since $\mathrm{x} \in \mathfrak{F}\mathfrak{f}$, it follows by (4) that Φ is in $\mathfrak{M}\mathfrak{f}$. We also easily calculate that $Ind(3, \Phi) = 1$ (note that z is the third bound variable, $\mathrm{z} = \mathrm{v}_3$) and $Ind(1, \Phi) = 2$. The only

variables which occur in Φ are a, x, and z. $S(1, p, \Phi)$ is easily shown to be equal to

$$(\mu \mathrm{z})[(\mathrm{z} + p) \approx (p + 1)] \approx \mathrm{x}.$$

With the help of the notions introduced above we can express the fundamental postulate of the meta-system in the following way: *There exists a one-to-one mapping of expressions onto the set* $\mathfrak{Ff} \vee \mathfrak{Mf}$ *such that* (i) **1** *is mapped onto* 1, (ii) *the i-th free variable* $\mathbf{a}_i$ *is mapped onto* a_i, *and the i-th bound variable* $\mathbf{v}_i$ *is mapped onto* v_i, (iii) *if functional forms* F_1 *and* F_2 *are mapped onto integers* b_1 *and* b_2, *then* F_1 **sum** F_2, F_1 **mult** F_2 *and* F_1 **equ** F_2 *are mapped onto* $b_1 + b_2$, $b_1 \times b_2$, *and* $b_1 \approx b_2$, (iv) *if matrix-forms* M_1 *and* M_2 *are mapped onto integers* b_1 *and* b_2, *then* M_1 **imp** M_2 *is mapped onto* $b_1 \rightarrow b_2$, (v) *if a matrix-form* M *is mapped onto an integer* b, *then* (**min** $\mathbf{v}_j$) M *is mapped onto* $(\mu \mathrm{v}_j)b$.

3. Properties of functional forms and matrix-forms. We note here some simple theorems which follow easily from the inductive definition of the preceding section.

a. *If* v_h *or* a_h occurs *in* b, *then* $h < b$.
b. *If* v_h *does not* occur *in* b, *then* $Ind(h, b) = 1$.
c. *If* $p \in \mathfrak{Ff}$, *then* $S(i, p, b)$ *is in* $\mathfrak{Ff}$ *or in* $\mathfrak{Mf}$ *according as* b *is in* $\mathfrak{Ff}$ *or in* $\mathfrak{Mf}$.
d. $Ind(j, S(i, p, b)) \leqslant \max\,(Ind(j, p), Ind(j, b))$.
e. *If* a_i occurs *in* b, *then* $p \leqslant S(i, p, b)$.
f. *If* a_i *does not* occur *in* b, *then* $S(i, p, b) = b$.

The class $\mathfrak{Ex} = \mathfrak{Mf} \vee \mathfrak{Ff}$ will be called the class of expressions. A functional form b will be called a numerical expression if the index of each bound variable occurring in b is 1. The set of numerical expressions will be denoted by $\mathfrak{Ne}$:

$$b \in \mathfrak{Ne} \equiv b \in \mathfrak{Ff} \cdot (h)\,[Oc(\mathrm{v}_h, b) = 2 \supset Ind(h, b) = 1].$$

In view of a. and b. the definition of the set $\mathfrak{Ne}$ can also be expressed thus:

$$b \in \mathfrak{Ne} \equiv b \in \mathfrak{Ff} \cdot (h)_b\,[Ind(h, b) = 1].$$

A matrix-form all the bound variables of which have the index 1 will be called a matrix; their set will be denoted by $\mathfrak{M}$:

$$b \in \mathfrak{M} \equiv b \in \mathfrak{Mf} \cdot (h)_b\,[Ind(h, b) = 1].$$

From d. and the definitions of sets $\mathfrak{Ne}$ and $\mathfrak{M}$ we infer easily

g. *If p is a numerical expression and b is in $\mathfrak{Ne}$ or in $\mathfrak{M}$, then so is $S(i, p, b)$.*

Theorem g. says that the operation of substitution leads from matrices to matrices and from numerical expressions to numerical expressions provided that the "substituend" p be a numerical expression.

A matrix without free variables is called a sentence. The set of sentences will be denoted by $\mathfrak{S}$:

$$b \in \mathfrak{S} \equiv b \in \mathfrak{M} \cdot (h)_b\,[Oc(\mathrm{a}_h, b) = 1].$$

A numerical expression without free variables is called a numeral and their set is denoted by $\mathfrak{Num}$:

$$b \in \mathfrak{Num} \equiv b \in \mathfrak{Ne} \cdot (h)_b\,[Oc(\mathrm{a}_h, b) = 1].$$

In particular the expressions

$$1,\ 1 \mathbin{\boldsymbol{+}} 1,\ (1 \mathbin{\boldsymbol{+}} 1) \mathbin{\boldsymbol{+}} 1,\ \ldots$$

are numerals. The n-th of them will be denoted by $D(n)$ or by D_n and called the n-th digit. The inductive definition of the function $D(n)$ is

$$D(1) = 1,\ D(n + 1) = D(n) \mathbin{\boldsymbol{+}} 1.$$

We prove now some theorems concerning iterations of the function S.

h. *If $b \in \mathfrak{Ex}$ and $p, q \in \mathfrak{Ff}$, then*

$$S(i, p, S(j, q, b)) = \begin{cases} S(j, S(i, p, q), S(i, p, b)) & \text{if } i \neq j \\ S(i, S(i, p, q), b) & \text{if } i = j. \end{cases}$$

Proof. The theorem is easily verified when $b = 1$ or b is a variable. Since the operation S is distributive with respect to

the operations $\dotplus$, $\times$, $\approx$, $\rightarrow$, and $(\mu \mathrm{v}_h)$, we infer by an easy induction that the theorem holds for an arbitrary b in $\mathfrak{Er}$.

i. *If* a_h *(or* v_h*) occurs neither in* p *nor in* b*, then it does not occur in* $S(i, p, b)$.

Dealing with repeated substitutions, it is convenient to use the notation $S_p^i\, b$ instead of $S(i, p, b)$. The formulas of theorem h. take then on the form

$$S_p^i\, S_q^j\, b = S_{S(i,p,q)}^j\, S_p^i\, b \quad \textit{if} \quad i \neq j,$$
$$S_p^i\, S_q^i\, b = S_{S(i,p,q)}^i\, b.$$

In particular, we obtain

j. *If* $i \neq j$ *and* $S(i, p, q) = q$*, then the operations* S_p^i *and* S_q^j *are commutative.*

From this theorem and f. we obtain

k. *If* p *and* q *are numerals, then the operations* S_p^i *and* S_q^j *are commutative* $(j \neq i)$.

In Chapter IV we shall use the following abbreviated notation for several substitutions performed one after the other: If $\mathrm{a}_i, \mathrm{a}_j, \ldots, \mathrm{a}_m$ are all the free variables which occur in an expression Ω $(i < j < \ldots < m)$, then we put $\Omega(\varphi, \psi, \ldots, \vartheta) = S_\varphi^i\, S_\psi^j \ldots S_\vartheta^m\, \Omega$. In particular, we can write $\Omega(\mathrm{a}_i, \mathrm{a}_j, \ldots, \mathrm{a}_m)$ for Ω itself which closely resembles the customary notation of the functional calculus.

Let π be a permutation of indices $i, j, \ldots, m$, and denote by $\pi(i), \pi(j), \ldots, \pi(m)$ the integers into which $i, j, \ldots, m$ are carried by π. Further put

$$\Omega' = \Omega(\mathrm{a}_{m+\pi(i)}, \mathrm{a}_{m+\pi(j)}, \ldots, \mathrm{a}_{m+\pi(m)}).$$

We shall prove the following theorem

l. *If* $\varphi_i, \varphi_j, \ldots, \varphi_m$ *are numerals, then*

$$\Omega'(\varphi_i, \varphi_j, \ldots, \varphi_m) = \Omega(\varphi_{\pi(i)}, \varphi_{\pi(j)}, \ldots, \varphi_{\pi(m)}).$$

Proof. Write $\overline{S}_j^i \Phi$ for $S(i, \mathrm{a}_j, \Phi)$ and $\overline{\overline{S}}_j^i \Phi$ for $S(i, \varphi_j, \Phi)$. The left side of the desired equation is

$$\overline{\overline{S}}_i^{m+i}\, \overline{\overline{S}}_j^{m+j} \ldots \overline{\overline{S}}_m^{m+m}\, \Omega' = \overline{\overline{S}}_i^{m+i} \ldots \overline{\overline{S}}_m^{m+m}\, \overline{S}_{m+\pi(i)}^i \ldots \overline{S}_{m+\pi(m)}^m\, \Omega.$$

It follows easily from k. that all substitutions occurring in this equation are commutative with each other. Since

$$\overline{\overline{S}}{}^{m+\pi(i)}_{\pi(i)} \, \overline{S}{}^{i}_{m+\pi(i)} = \overline{\overline{S}}{}^{i}_{\pi(i)}$$

(cf. h.), we obtain

$$\overline{\overline{S}}{}^{m+i}_{i} \, \overline{\overline{S}}{}^{m+j}_{j} \ldots \overline{\overline{S}}{}^{m+m}_{m} \, \Omega' = \overline{\overline{S}}{}^{i}_{\pi(i)} \, \overline{\overline{S}}{}^{j}_{\pi(j)} \ldots \overline{\overline{S}}{}^{m}_{\pi(m)} \, \Omega$$

which completes the proof of the theorem.

m. *Let* $b \in \mathfrak{Ex}$ *and let* $\mathrm{v}_h, \mathrm{v}_k, \ldots, \mathrm{v}_m$ *be all the* b o u n d v a r i a b l e s *which* o c c u r *in* b *and have there the index* 2. *Let* $\mathrm{a}_p, \mathrm{a}_q, \ldots, \mathrm{a}_r$ *be* f r e e v a r i a b l e s *which do not* o c c u r *in* b. *There exists then a uniquely determined* e x p r e s s i o n c *such that*

$$b = S^{p}_{\mathrm{v}_h} \, S^{q}_{\mathrm{v}_k} \ldots S^{r}_{\mathrm{v}_m} \, c$$

and such that no b o u n d v a r i a b l e o c c u r s *in* c *with the index* 2 (*hence* c *is either a* m a t r i x *or a* n u m e r i c a l e x p r e s s i o n) [7].

4. Explicit definition of sets $\mathfrak{Ff}$, $\mathfrak{Mf}$, and of functions Oc, Ind, and S. In this section we shall transform the inductive definition given in section **2** into an explicit one. The method is in principle the same as that which we used in Chapter I, section **5** but is a little more complicated because we are dealing here with a simultaneous definition of two sets and three functions.

First we introduce eight auxiliary relations which are closely connected with the inductive clauses (1)—(6) of section **2**.

$\mathfrak{R}_1(j', j'', k, l, m, a, i, n, p) \equiv (j' = j'' = k = l = m = 1)$.

$\mathfrak{R}_2(j', j'', k, l, m, a, i, n, p) \equiv (j' \in \mathfrak{Vf}) \cdot (j'' = 1) \cdot$
$[(k = 1) \cdot (a \neq j') \vee (k = 2) \cdot (a = j')] \cdot (l = 1) \cdot$
$[(m = j') \cdot (\mathrm{a}_n \neq j') \vee (m = p) \cdot (\mathrm{a}_n = j')]$.

$\mathfrak{R}_3(j', j'', k, l, m, a, i, n, p) \equiv (j' \in \mathfrak{Vb}) \cdot (j'' = 1) \cdot$
$[(k = 1) \cdot (a \neq j') \vee (k = 2) \cdot (a = j')] \cdot [(l = 2) \cdot (\mathrm{v}_i = j') \vee$
$(l = 1) \cdot (\mathrm{v}_i \neq j')] \cdot (m = j')$.

[7] The expression c corresponds to what Hilbert and Bernays [11], Vol. 1, p. 89 call the "Nennform" of an expression.

To explain these definitions let us suppose that a, i, n, p are fixed integers. a is to be thought of as a variable (free or bound) i as the subscript in the variable v_i, n as the subscript in the variable a_n, and p as a functional form. Suppose that one of the formulas $\mathfrak{R}_s(j', j'', \ldots, p)$ holds ($s = 1, 2, 3$). j' is then either 1 or a free variable or a bound variable; k is 2 or 1 depending on whether a occurs in j' or not, l is equal to $Ind(i, j')$, and m to $S(n, p, j')$. The role of j'' which was put equal to 1 in all the cases will be explained later.

$\mathfrak{R}_4\,(j_1', j_2', j_3', j_1'', j_2'', j_3'', k_1, k_2, k_3, l_1, l_2, l_3, m_1, m_2, m_3)$
$\equiv (j_1'' = j_2'' = 1)\cdot(j_3'' = 1)\cdot(j_3' = j_1' \boldsymbol{+} j_2')\cdot$
$(k_3 = \max\,(k_1, k_2))\cdot(l_3 = \max\,(l_1, l_2))\cdot(m_3 = m_1 \boldsymbol{+} m_2)$.

$\mathfrak{R}_5$ similar to $\mathfrak{R}_4$ but with the sign $\boldsymbol{+}$ replaced by $\times$.

$\mathfrak{R}_6$ similar to $\mathfrak{R}_4$ but with the sign $\boldsymbol{+}$ replaced by $\approx$ and the equation $j_3'' = 1$ replaced by $j_3'' = 2$.

$\mathfrak{R}_7$ similar to $\mathfrak{R}_4$ but with the sign $\boldsymbol{+}$ replaced by $\rightarrow$ and the equations $(j_1'' = j_2'' = 1)\cdot(j_3'' = 1)$ replaced by $(j_1'' = j_2'' = 2)\cdot(j_3'' = 2)$.

$\mathfrak{R}_8\,(j_1', j_2', j_1'', j_2'', s, k_1, k_2, l_1, l_2, m_1, m_2, a, i) \equiv (j_1'' = 2)\cdot(s \in \mathfrak{Vb})\cdot$
$(j_2'' = 1)\cdot(j_2' = (\mu s)j_1')\cdot$
$\{[(a = s) \vee (k_1 = 2)]\cdot(k_2 = 2) \vee (a \neq s)\cdot(k_1 = 1)\cdot(k_2 = 1)\}\cdot$
$[(l_2 = 1)\cdot(\mathrm{v}_i = s) \vee (l_2 = l_1)\cdot(\mathrm{v}_i \neq s)]\cdot(m_2 = (\mu s)m_1)$.

To explain these definitions assume that j_1' and j_2' are functional forms, k_t is 2 or 1 depending on whether a occurs or does not occur in j_t', l_t is equal to $Ind(i, j_t')$, and m_t to $S(n, p, j_t')$ ($t = 1, 2$). If the relation $\mathfrak{R}_4(j_1', j_2', \ldots, m_3)$ holds, then $j_3' = j_1' \boldsymbol{+} j_2'$, k_3 is 2 or 1 depending on whether a occurs or does not occur in j_3', $l_3 = Ind(i, j_3')$, and $m_3 = S(n, p, j_3')$. The meaning of relations $\mathfrak{R}_5$ — $\mathfrak{R}_7$ is similar. The role of the integers j_t'' is this: If j_t' is a functional form, then j_t'' is 1 and if j_t' is a matrix-form, then j_t'' is 2. Hence, the value of j_t'' enables us to distinguish between the elements of $\mathfrak{Ff}$ and those of $\mathfrak{Mf}$.

The meaning of the relation $\mathfrak{R}_8$ is similar: it corresponds to the process of forming the functional form $(\mu s)j_1'$ from the matrix-form j_1' and the bound variable s. Accordingly, we require that the value of j_1'' be 2 and that of j_2'' be 1.

Let now a, i, n, p be arbitrary integers. We shall show that the condition

$$(1) \qquad x \in \mathfrak{Ex}$$

implies the existence of an integer $w < x$ and of five sequences

$$(2) \qquad j_v', j_v'', k_v, l_v, m_v \quad (v = 1, 2, \ldots, w)$$

such that the following conditions are satisfied:

$$(3) \qquad j_v' \leqslant x,\ j_v'' \leqslant 2,\ k_v \leqslant 2,\ l_v \leqslant 2,\ m_v \leqslant 10^{5^v} \cdot (p + x)^{4^v} \qquad (v = 1, 2, \ldots, w),$$

$$(4) \qquad j_w' = x,\ k_w = Oc(a, x),\ l_w = Ind(i, x),\ m_w = S(n, p, x),$$

(5) *for every* $v \leqslant w$ *either*

(5_1) *there is an* $s \leqslant 3$ *such that*

$$\mathfrak{R}_s(j_v', j_v'', k_v, l_v, m_v, a, i, n, p)$$

or

(5_2) *there are* ϱ, σ, *and* s *such that* $\varrho < v$, $\sigma < v$, $s < 5$ *and either*

$$\mathfrak{R}_{s+3}(j_\varrho', j_\sigma', j_v', j_\varrho'', j_\sigma'', j_v'', k_\varrho, k_\sigma, k_v, l_\varrho, l_\sigma, l_v, m_\varrho, m_\sigma, m_v)$$

or

$$\mathfrak{R}_8(j_\varrho', j_v', j_\varrho'', j_v'', j_\sigma', k_\varrho, k_v, l_\varrho, l_v, m_\varrho, m_v, a, i).$$

This will be shown as follows. Assume that (1) holds. If $x = 1$ or $x \in \mathfrak{Vf}$ or $x \in \mathfrak{Vb}$, then it is sufficient to take $w = 1$, $j_1'' = 1$, $k_1 = Oc(a, x)$, $l_1 = Ind(i, x)$, $m_1 = S(n, p, x)$.

Assume now that x_1 and x_2 are in $\mathfrak{Ex}$ and that there exist sequences (with $w(1)$ and $w(2)$ terms) satisfying conditions (3), (4), and (5) (of course with x replaced by x_1 or x_2 respectively). Let x be one of the expressions $x_1 + x_2$, $x_1 \times x_2$, $x_1 \approx x_2$, $x_1 \rightarrow x_2$, $(\mu x_1)x_2$, and consider five sequences obtained from the sequences corresponding

to x_1 and x_2 by pulling them together into single sequences with $w(1) + w(2)$ terms. We add to the first sequence the number x as the last term of the sequence. To the second sequence we add similarly the number 1 or 2 depending on whether x is in $\mathfrak{Fk}$ or in $\mathfrak{Mk}$. Finally, we add $Oc(a, x)$, $Ind(i, x)$, and $S(n, p, x)$ as the last terms to the third, fourth, and fifth sequence. In this way we obtain five sequences with $w(1) + w(2) + 1$ terms, and we see easily that these sequences satisfy conditions (4) and (5). The first four inequalities (3) are evidently satisfied. Thus it remains to show that the last of the inequalities (3) is also satisfied and that $w(1) + w(2) + 1$ does not exceed x.

Let us put $w = w(1) + w(2) + 1$. According to the assumptions concerning the sequences corresponding to x_1 and x_2 we have for $v \leqslant w - 1$

$$m_v \leqslant 10^{5^{w(1)}} \cdot (p + x_1)^{4^{w(1)}}$$

or

$$m_v \leqslant 10^{5^{w(2)}} \cdot (p + x_2)^{4^{w(2)}}$$

and hence

$$m_v \leqslant 10^{5^{w-1}} \cdot (p + x)^{4^{w-1}}.$$

If $v = w$, m_v has one of the values $m_{w(1)} + m_{w(2)}$, $m_{w(1)} \times m_{w(2)}$, $m_{w(1)} \approx m_{w(2)}$, $m_{w(1)} \rightarrow m_{w(2)}$, $(\mu j'_{w(1)}) m_{w(2)}$ and hence m_w does not exceed the value

$$4 . J_3(7, \max(m_{w(1)}, j'_{w(1)}), m_{w(2)}) + 3.$$

Since $j'_{w(1)} \leqslant x$, $\max(m_{w(1)}, j'_{w(1)}) \leqslant 10^{5^{w-1}} \cdot (p + x)^{4^{w-1}}$, and we obtain, using the inequality for the function J_3 established at the end of Chapter I, section 1, p. 15

$$\begin{aligned} m_w &\leqslant 4 . J_3(7, 10^{5^{w-1}} \cdot (p + x)^{4w-1}, 10^{5^{w-1}} \cdot (p + x)^{4^{w-1}}) + 3 \\ &\leqslant 4 \cdot [7 + 10^{5^{w-1}} \cdot (p + x)^{4w-1} + 10^{5^{w-1}} \cdot (p + x)^{4^{w-1}}]^4 + 3 \\ &< 4 . 81 . 10^{5^{w-1} . 4} \cdot (p + x)^{4^{w-1} . 4} + 3 \\ &< 400 . 10^{5^{w-1} . 4} \cdot (p + x)^{4^w} < 10^{5^w} \cdot (p + x)^{4^w}. \end{aligned}$$

By a similar evaluation we can also prove that $w = w(1) + w(2) + 1$ does not exceed x.

We have thus completed the proof that (1) implies the existence of sequences (2) with the properties (3) — (5).

Conversely, let us assume that there are sequences (2) satisfying the conditions (3), (4), and (5). By an easy induction on v we show that this assumption implies that

$$j'_v \in \mathfrak{Ex},\ k_v = Oc(a, j'_v),\ l_v = Ind(i, j'_v),\ m_v = S(n, p, j'_v),$$
$$(j''_v = 1) \equiv (j'_v \in \mathfrak{Ff}), \qquad (j''_v = 2) \equiv (j'_v \in \mathfrak{Mf}).$$

Let now e', e'', f, g, h be integers which represent the sequences (2). From the inequalities (3) we obtain with the help of theorem 3 (Chapter I, section 2, p. 17) the following inequalities:

$$(6) \qquad \begin{cases} e' < T(x),\ e'' < T(x+2),\ f < T(x+2), \\ g < T(x+2),\ h < T(F(p, x)) \end{cases}$$

where

$$F(p, x) = 10^{5^x} \cdot (p + x)^{4^x}.$$

From (3), (4), and (5) we infer that the integers e', e'', f, g, h satisfy the relation

$$\mathfrak{R}(e', e'', f, g, h, x, a, i, n, p) \equiv [L(e') = L(e'') = L(f) = L(g) = L(h)] \cdot (\bar{e}'_* = x).$$
$$(\nu)_{L(e')} \{(\exists s)_3\, \mathfrak{R}_s(\bar{e}'_\nu, \bar{e}''_\nu, \bar{f}_\nu, \bar{g}_\nu, \bar{h}_\nu, a, i, n, p) \vee (\exists \varrho)_\nu (\exists \sigma)_\nu (\exists s)_5 [\mathfrak{R}_{s+3}(\bar{e}'_\varrho, \bar{e}'_\sigma, \bar{e}'_\nu, \bar{e}''_\varrho, \bar{e}''_\sigma, \bar{e}''_\nu, \bar{f}_\varrho, \bar{f}_\sigma, \bar{f}_\nu, \bar{g}_\varrho, \bar{g}_\sigma, \bar{g}_\nu, \bar{h}_\varrho, \bar{h}_\sigma, \bar{h}_\nu) \vee \mathfrak{R}_8(\bar{e}'_\varrho, \bar{e}'_\nu, \bar{e}''_\varrho, \bar{e}''_\nu, \bar{e}'_\sigma, \bar{f}_\varrho, \bar{f}_\nu, \bar{g}_\varrho, \bar{g}_\nu, \bar{h}_\varrho, \bar{h}_\nu, a, i)]\}.$$

Hence if $x \in \mathfrak{Ex}$, there are integers $e', e'', \ldots, h$ satisfying the inequalities (6), and such that

$$\mathfrak{R}(e', e'', f, g, h, x, a, i, n, p).$$

Conversely, if such integers exist, then the sequences represented by these integers satisfy the conditions (3), (4), and (5) and hence

for their last terms we obtain the relations

$$\bar{e}'_* = x \in \mathfrak{Ex},\ \bar{f}_* = Oc(a, x),\ \bar{g}_* = Ind(i, x),$$
$$\bar{h}_* = S(n, p, x),$$
$$(\bar{e}''_* = 1) \equiv (x \in \mathfrak{Ff}),\ (\bar{e}''_* = 2) \equiv (x \in \mathfrak{Mf}).$$

The result which we have thus obtained can be summarized in the following equivalence:

$$(7')\left\{\begin{array}{l}(x \in \mathfrak{Ff})\cdot(y = Oc(a, x))\cdot(z = Ind(i, x))\cdot(t = S(n, p, x)) \equiv \\ \equiv (\exists e')_{T(x)}\,(\exists e'')_{T(x+2)}\,(\exists f)_{T(x+2)}\,(\exists g)_{T(x+2)}\,(\exists h)_{T(F(p,x))} \\ \mathfrak{R}(e', e'', f, g, h, x, a, i, n, p)\cdot(x = \bar{e}'_*)\cdot(1 = \bar{e}''_*)\cdot(y = \bar{f}_*)\cdot \\ (z = \bar{g}_*)\cdot(t = \bar{h}_*)\end{array}\right.$$

and in a similar equivalence (7″) for $x \in \mathfrak{Mf}$ where the factor $1 = \bar{e}''_*$ on the right hand side has to be replaced by $2 = \bar{e}''_*$.

These explicit definitions could have been adopted instead of the inductive definition given in section 2 on p. 28. The relations (1) — (6) which we adopted there as definitions can now be deduced from (7′) and (7″). We omit the details of this deduction because it will not be needed in the sequel.

5. Axioms of (S). A *matrix* Ω will be called an *axiom* of the propositional calculus (in symbols $\Omega \in \mathfrak{Ax}_{\mathrm{I}}$) if there are *matrices* Φ, Ψ, Θ such that $\Omega = (\Phi \to \Psi) \to [(\Psi \to \Theta) \to (\Phi \to \Theta)]$ or $\Omega = [(\Phi \to \Psi) \to \Phi] \to \Phi$ or $\Omega = \Phi \to (\Psi \to \Phi)$ or $\Omega = (1 \approx \approx 1 + 1) \to \Phi$ [8].

Observing that if Φ, Ψ, and Θ satisfy any of these equations, they are necessarily less than Ω, we can formulate the definition of the set $\mathfrak{Ax}_{\mathrm{I}}$ in the following way:

$$\Omega \in \mathfrak{Ax}_{\mathrm{I}} \equiv (\exists\Phi)_\Omega(\exists\Psi)_\Omega(\exists\Theta)_\Omega[(\Phi \in \mathfrak{M})\cdot(\Psi \in \mathfrak{M})\cdot(\Theta \in \mathfrak{M})\cdot \\ \cdot(\{\Omega = (\Phi \to \Psi) \to [(\Psi \to \Theta) \to (\Phi \to \Theta)]\} \vee \{\Omega = [(\Phi \to \Psi) \to \Phi] \to \\ \to \Phi\} \vee [\Omega = \Phi \to (\Psi \to \Phi)] \vee [\Omega = (1 \approx 1 + 1) \to \Phi])].$$

[8] These axioms for the propositional calculus have been given by Church [3]. We replaced his propositional constant f by $1 \approx 1 + 1$.

A matrix Ω is called an axiom of the functional calculus (in symbols $\Omega \in \mathfrak{Ax}_{\mathrm{II}}$) if one of the following conditions is satisfied: (i) there is a matrix Φ, a bound variable v_h which does not occur in Φ, and a free variable a_i which occurs in Φ such that either

$$(1) \qquad \Omega = \Phi \rightarrow S(i, (\mu \mathrm{v}_h) S(i, \mathrm{v}_h, \Phi), \Phi)$$

or

$$(2) \qquad \Omega = S(i, (\mu \mathrm{v}_h)\, S(i, \mathrm{v}_h, \Phi), \Phi \rightarrow (1 \approx 1 + 1)) \rightarrow (\mu \mathrm{v}_h) S(i, \mathrm{v}_h, \Phi) \approx 1,$$

(ii) there is a matrix Φ, a free variable a_i which occurs in Φ, and three variables a_k, a_l, v_h which do not occur in Φ such that

$$(3) \qquad \Omega = [(\mu \mathrm{v}_h) S(i, \mathrm{v}_h, \Phi) \approx (\mathrm{a}_k + \mathrm{a}_l)] \rightarrow [S(i, \mathrm{a}_k, \Phi) \rightarrow (1 \approx 1 + 1)].$$

A more formal definition of $\mathfrak{Ax}_{\mathrm{II}}$ is

$$\begin{aligned} &\Omega \in \mathfrak{Ax}_{\mathrm{II}} \equiv (\exists \Phi)_\Omega\, (\exists i)_\Phi\, (\exists h)_\Omega\, \{(\Phi \in \mathfrak{M}) \cdot (Oc(\mathrm{a}_i, \Phi) = 2) \cdot \\ &(Oc(\mathrm{v}_h, \Phi) = 1) \cdot [(\ldots.) \vee (\ldots.)]\} \vee \\ &(\exists \Phi)_\Omega\, (\exists i)_\Phi\, (\exists h)_\Omega\, (\exists k)_\Omega\, (\exists l)_\Omega\, [(\Phi \in \mathfrak{M}) \cdot (Oc(\mathrm{a}_i, \Phi) = 2). \\ &(Oc(\mathrm{v}_h, \Phi) = 1) \cdot (Oc(\mathrm{a}_k, \Phi) = 1) \cdot (Oc(\mathrm{a}_l, \Phi) = 1) \cdot (\ldots.)] \end{aligned}$$

where the blanks "...." are to be filled by the equations (1), (2), and (3).

The set $\mathfrak{Ax}_{\mathrm{III}}$ of the axioms of identity is defined as follows:

$$\begin{aligned} &\Omega \in \mathfrak{Ax}_{\mathrm{III}} \equiv \Omega = (\mathrm{a} \approx \mathrm{a}) \vee \Omega = (\mathrm{a} \approx \mathrm{b}) \rightarrow (\mathrm{b} \approx \mathrm{a}) \vee \\ &\Omega = (\mathrm{a} \approx \mathrm{b}) \rightarrow [(\mathrm{b} \approx \mathrm{c}) \rightarrow (\mathrm{a} \approx \mathrm{c})] \vee \Omega = (\mathrm{a} \approx \mathrm{b}) \rightarrow (\mathrm{a} + \mathrm{c} \approx \mathrm{b} + \mathrm{c}) \\ &\vee \Omega = (\mathrm{a} \approx \mathrm{b}) \rightarrow (\mathrm{a} \times \mathrm{c} \approx \mathrm{b} \times \mathrm{c}) \vee (\exists \Phi)_\Omega (\exists i)_\Phi (\exists j)_\Omega (\exists k)_\Omega \\ &\{(\Phi \in \mathfrak{M}) \cdot (Oc(\mathrm{a}_i, \Phi) = 2) \cdot (Oc(\mathrm{v}_j, \Phi) = Oc(\mathrm{v}_k, \Phi) = 1) \cdot \\ &[\Omega = (\mu \mathrm{v}_j) S(i, \mathrm{v}_j \Phi) \approx (\mu \mathrm{v}_k) S(i, \mathrm{v}_k, \Phi)]\} \vee \\ &(\exists \Phi)_\Omega (\exists \Psi)_\Omega (\exists i)_\Omega (\exists j)_\Omega [(\Phi \in \mathfrak{M}) \cdot (\Psi \in \mathfrak{M}) \cdot \{\Omega = \\ &(\Phi \rightarrow \Psi) \rightarrow [(\Psi \rightarrow \Phi) \rightarrow (\mu \mathrm{v}_j) S(i, \mathrm{v}_j, \Phi) \approx (\mu \mathrm{v}_j) S(i, \mathrm{v}_j, \Psi)]\}]. \end{aligned}$$

Finally, we define the set $\mathfrak{Ax}_{\mathrm{IV}}$ of the axioms of arithmetic:

$$\begin{aligned} \Omega \in \mathfrak{Ax}_{\mathrm{IV}} \equiv\ & \Omega = (\mathrm{a} + \mathrm{b} \approx 1) \rightarrow (1 \approx 1 + 1) \vee \\ & \Omega = [(\mathrm{a} \approx 1) \rightarrow (1 \approx 1 + 1)] \rightarrow (\mu \mathrm{x})(\mathrm{x} + 1 \approx \mathrm{a}) + 1 \approx \mathrm{a} \end{aligned}$$

$$\mathbf{v}\,\Omega = \mathrm{a} + (\mathrm{b} + \mathrm{c}) \approx (\mathrm{a} + \mathrm{b}) + \mathrm{c}$$
$$\mathbf{v}\,\Omega = (\mathrm{a} + \mathrm{c} \approx \mathrm{b} + \mathrm{c}) \rightarrow (\mathrm{a} \approx \mathrm{b})$$
$$\mathbf{v}\,\Omega = (\mathrm{a} + \mathrm{b} \approx \mathrm{b} + \mathrm{a})$$
$$\mathbf{v}\,\Omega = \mathrm{a} \times 1 \approx \mathrm{a}$$
$$\mathbf{v}\,\Omega = \mathrm{a} \times (b + \mathrm{c}) \approx (\mathrm{a} \times \mathrm{b}) + (\mathrm{a} \times \mathrm{c})$$
$$\mathbf{v}\,\Omega = \mathrm{a} \times \mathrm{b} \approx \mathrm{b} \times \mathrm{a}.$$

The set $\mathfrak{Ax}$ of all axioms can now be defined thus:

$$\Omega \in \mathfrak{Ax} \equiv \Omega \in \mathfrak{Ax}_{\mathrm{I}} \,\mathbf{v}\, \Omega \in \mathfrak{Ax}_{\mathrm{II}} \,\mathbf{v}\, \Omega \in \mathfrak{Ax}_{\mathrm{III}} \,\mathbf{v}\, \Omega \in \mathfrak{Ax}_{\mathrm{IV}}.$$

The axioms are for the most part self explanatory. Axioms of the group I give a sufficient basis for the propositional calculus. The first axiom of the group II has the following intuitive content: If there is an integer satisfying the condition expressed by the matrix Φ, then the integer denoted by the numerical expression $(\mu \mathrm{v}_j)S(i, \mathrm{v}_j, \Phi)$ satisfies the same condition. The second axiom of the group II states that if the integer denoted by $(\mu \mathrm{v}_j)S(i, \mathrm{v}_j, \Phi)$ is not 1, then it satisfies the condition expressed by Φ. Finally, the third axiom of the group II states that if the integer denoted by $(\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi)$ is the sum of two integers, then none of these integers satisfies the condition expressed by Φ. The three axioms taken together characterize the integer denoted by $(\mu \mathrm{v}_j)S(i, \mathrm{v}_j, \Phi)$ as the smallest integer satisfying the condition expressed by Φ or 1 if no such integer exists.

The various clauses concerning the bound variables have been added in order to avoid collisions between the variables. We show by an example that the first axiom of the group II would be intuitively false in some cases, had the clause that v_h does not occur in Φ been omitted from the formulation of the axiom: Let $\Phi = (\mu \mathrm{x})\,(\mathrm{x} + \mathrm{a} \approx D_6) \approx D_4$; the condition expressed by Φ is satisfied exclusively by the integer 2 and $(\mu \mathrm{y})[(\mu \mathrm{x})(\mathrm{x} + \mathrm{y} \approx D_6) \approx D_4]$ denotes 2 but $(\mu \mathrm{x})[(\mu \mathrm{x})(\mathrm{x} + \mathrm{x} \approx D_6) \approx D_4]$ denotes 1.

The last but one axiom of the group III gives the possibility of changing the bound variable under the sign μ in cases when no collision of variables occurs. The last axiom of this group states that if two conditions are equivalent, then the smallest integers satisfying these conditions are equal.

6. Formal proofs. We denote by Mp a function such that

$$Mp(\Phi \to \Psi, \Phi) = \Psi$$

and $Mp(\Omega, \Phi) = \Omega$ if Ω has not the form $\Phi \to \Psi$. The symbol Mp is an abbreviation of the words "modus ponens".

An explicit definition of Mp is

$$Mp = \min_3 \mathfrak{R}$$

where

$$\mathfrak{R} = \lambda xyz[(y \to z = z) \vee (z = x)].$$

Indeed, if $\Omega = \Phi \to \Psi$, then the smallest z such that $\mathfrak{R}(\Omega, \Phi, z)$ is Ψ; if Ω has not the form $\Phi \to \Psi$, then the smallest such z is Ω.

Observe that the relation $\mathfrak{R}$ satisfies the condition

$$(x)(y)(\exists z)\mathfrak{R}(x, y, z).$$

Let $\mathfrak{K}$ be an arbitrary class of matrices. A finite sequence of matrices

(1) $$\Phi_1, \Phi_2, \ldots, \Phi_n$$

will be called a *formal* $\mathfrak{K}$*-proof* if for every $j \leqslant n$ one of the following conditions is satisfied:

(2) $\Phi_j \in \mathfrak{K}$, (3) $\Phi_j \in \mathfrak{A}\mathfrak{x}$,

(4) *there are k and l such that $k < j$, $l < j$, and $\Phi_j = Mp(\Phi_k, \Phi_l)$,*

(5) *there is a $k < j$, a variable a_h, and a numerical expression φ such that $\Phi_j = S(h, \varphi, \Phi_k)$.*

Observe that h in (5) is necessarily less than Φ_k, and φ less than Φ_j (cf. section 3, theorems a. and e., p. 29).

A matrix Φ is called $\mathfrak{K}$*-provable* if it is the last term of a formal $\mathfrak{K}$-proof. We write then $\Phi \in \mathfrak{T}_{\mathfrak{K}}$ or $\mathfrak{K}\text{-}\vdash\Phi$. If $\mathfrak{K}$ is void, we use the terms *formal proof*, *provable* and write $\Phi \in \mathfrak{T}$ or $\vdash\Phi$.

We shall denote by $\mathfrak{P}_{\mathfrak{K}}$ the set of integers which represent formal $\mathfrak{K}$-proofs. The exact definition of this class is the following:

$$\begin{aligned}&g \in \mathfrak{P}_{\mathfrak{K}} \equiv (j)_{L(g)}((\bar{g}_j \in \mathfrak{M}). \{(\bar{g}_j \in \mathfrak{K}) \vee (\bar{g}_j \in \mathfrak{A}\mathfrak{x}) \vee \\ &(\exists k)_j (\exists l)_j (\bar{g}_j = Mp(\bar{g}_k, \bar{g}_l)) \vee \\ &(\exists k)_j (\exists \varphi)_{\bar{g}_j} (\exists h)_{\bar{g}_k} [(\varphi \in \mathfrak{N}\mathrm{e}) \cdot (\bar{g}_j = S(h, \varphi, \bar{g}_k))]\}).\end{aligned}$$

The set $\mathfrak{T}_{\mathfrak{R}}$ can now be defined by the equivalence

$$\Phi \in \mathfrak{T}_{\mathfrak{R}} \equiv (\exists g)[(g \in \mathfrak{P}_{\mathfrak{R}}) \cdot (\bar{g}_{*} = \Phi)].$$

The existential quantifier on the right hand side of this equivalence is not bounded. This seemingly unimportant remark will essentially influence the proofs given in Chapter VI (cf. also Introduction, p. 11).

Numerous other syntactical notions are definable in terms of the notions introduced thus far. The most important are the following.

A set $\mathfrak{R}$ of matrices is called *closed* if $\mathfrak{R} = \mathfrak{T}_{\mathfrak{R}}$. It is called *consistent* if there are matrices which are not in $\mathfrak{T}_{\mathfrak{R}}$. A sentence Θ is called *undecidable* with respect to $\mathfrak{R}$ if neither Θ nor

$$\Theta \rightarrow (1 \approx 1 + 1)$$

are in $\mathfrak{T}_{\mathfrak{R}}$. If there are no sentences undecidable with respect to $\mathfrak{R}$, then $\mathfrak{R}$ is called *complete.*

We note the following well-known theorem:

Theorem 1. *In order that $\mathfrak{R}$ be consistent, it is necessary and sufficient that for every sentence Θ either Θ or $\Theta \rightarrow (1 \approx 1 + 1)$ be not $\mathfrak{R}$-provable. In order that $\mathfrak{R}$ be complete it is necessary and sufficient that for every sentence Θ either Θ or $\Theta \rightarrow (1 \approx 1 + 1)$ be $\mathfrak{R}$-provable. If Θ is undecidable, then the class consisting of Θ alone is consistent* [9].

The notion of consistency can be considered as a first approximation to the vague notion of "intuitive truth" of arithmetical sentences. An inconsistent closed set $\mathfrak{R}$ contains certainly sentences which are intuitively false although the converse of this statement does not hold in general.

A further important notion is that of *ω-consistency* [10]: A set $\mathfrak{R}$ of matrices is called *ω-consistent* if there is no matrix Φ with

[9] Cf. Tarski [20], pp. 9, 27, 31.

[10] This notion has been first introduced by Gödel [9], p. 187.

the following properties:

$$\mathrm{a}_1 \textit{ but no other free variable occurs in } \Phi,$$
$$(n)\, S(1, D_n, \Phi \to (1 \approx 1 + 1)) \in \mathfrak{T}_{\mathfrak{R}},$$
$$S(1, (\mu \mathrm{v}_h)\, S(1, \mathrm{v}_h, \Phi), \Phi) \in \mathfrak{T}_{\mathfrak{R}},$$

(v_h is an arbitrary bound variable which does not occur in Φ).

An ω-inconsistent closed set $\mathfrak{R}$ contains sentences which are intuitively false. Indeed, if all the sentences $S(1, D_n, \Phi \to 1 \approx 1 + 1)$ are intuitively true, then the sentence $S(1, (\mu \mathrm{v}_h) S(1, \mathrm{v}_h, \Phi), \Phi)$ is intuitively false. Hence the notion of ω-consistency can be considered as a second approximation to the vague notion of "truth". We shall see later that it is still a very imperfect approximation.

The relation of ordinary consistency to the ω-consistency is elucidated by the following theorem:

Theorem 2. *Every ω-consistent set is consistent.*

Proof. Assume that $\mathfrak{R}$ is an ω-consistent set and let Φ be an arbitrary matrix in which exactly one free variable a_1 occurs. If there is an n such that $S(1, D_n, \Phi \to (1 \approx 1 + 1))$ is not in $\mathfrak{T}_{\mathfrak{R}}$, then $\mathfrak{R}$ is consistent. Otherwise, the sentence $S(1, (\mu \mathrm{v}_h) S(1, \mathrm{v}_h, \Phi), \Phi)$ is not in $\mathfrak{T}_{\mathfrak{R}}$ and hence we infer again that the set $\mathfrak{R}$ is consistent. Hence $\mathfrak{R}$ is consistent in all cases.

We shall see in Chapter VI, section 4 that there are consistent but ω-inconsistent sets $\mathfrak{R}$. Hence the theorem converse to theorem 2 is false.

III

ARITHMETICAL THEOREMS PROVABLE IN (S)

A justification of the study of the system (S) can be seen in the fact that most theorems of the intuitive arithmetic can be written down by means of the symbols of (S) and become then provable matrices of (S). In other words, it is possible to make certain provable matrices of (S) correspond to theorems of intuitive arithmetic. We shall deal more extensively with the nature of this correspondence in Chapter V. The present Chapter is devoted to some examples of provable matrices which will be needed later.

1. **Abbreviations.** We shall use the following abbreviations:

(1) $\backsim \Phi = \Phi \rightarrow (1 \approx 1 + 1)$,

(2) $\Phi \vee \Psi = (\backsim \Phi) \rightarrow \Psi$,

(3) $\Phi \& \Psi = \backsim (\backsim \Phi \vee \backsim \Psi)$,

(4) $\Phi \leftrightarrow \Psi = (\Phi \rightarrow \Psi) \& (\Psi \rightarrow \Phi)$,

(5) $(\mathrm{E}\mathrm{v}_h)S(i, \mathrm{v}_h, \Phi) = S(i, (\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), \Phi)$, [1]

(6) $(\mathrm{A}\mathrm{v}_h)S(i, \mathrm{v}_h, \Phi) = \backsim (\mathrm{E}\mathrm{v}_h) \backsim S(i, \mathrm{v}_h, \Phi)$, [1]

(7) $\varphi \lessdot \psi = (\mathrm{E}\mathrm{v}_h)(\varphi + \mathrm{v}_h \approx \psi)$, [2]

(8) $\varphi - \psi = (\mu \mathrm{v}_h)(\mathrm{v}_h + \psi \approx \varphi)$. [2]

The definitions (1) — (4) introduce the connectives of the propositional calculus, the definitions (5) and (6) the existential and the general quantifier. Finally, (7) and (8) contain definitions of the arithmetical concepts of the inequality and of the difference.

2. **The propositional calculus.** The propositional calculus holds in (S) in the following sense: Let T be a formula of the propositional

[1] These definitions will be used only in cases when the variable v_h does not occur in Φ.

[2] v_h is a bound variable with the smallest index h such that v_h occurs neither in φ nor in ψ.

calculus. Replace in T the propositional variables by arbitrary matrices and the connectives *or*, *and*, *not* etc. by symbols $\vee$, &, $\backsim$ etc, and denote by Φ_T the matrix thus obtained. We have then the following theorem

Theorem 1. *If T is a tautology of the propositional calculus, then Φ_T is a provable matrix.*

We pass over an easy proof of this theorem. We shall often make use of it without quoting it explicitly.

3. The functional calculus. In this section we shall show that the functional calculus holds in (S). We begin with the following theorem:

Theorem 1. *If the variable* v_h *does not occur in the matrix* Φ, *then*

$$(1) \qquad \vdash \Phi \rightarrow (\mathrm{Ev}_h)S(i, \mathrm{v}_h, \Phi), \; \vdash (\mathrm{Av}_h)S(i, \mathrm{v}_h, \Phi) \rightarrow \Phi.$$

Proof. First formula results immediately from the axiom II 1 and the definition (5). Using tautologies of the propositional calculus, we obtain from

$$\vdash \backsim\Phi \rightarrow (\mathrm{Ev}_h)S(i, \mathrm{v}_h, \backsim\Phi)$$

the formula $\vdash \backsim (\mathrm{Ev}_h)S(i, \mathrm{v}_h, \backsim\Phi) \rightarrow \Phi$ which is identical with the second of the formulas (1).

Theorem 2. *If Φ and Ψ are matrices and* a_i *does not occur in Ψ, then from* $\vdash \Phi \rightarrow \Psi$ *follows* $\vdash (\mathrm{Ev}_h)S(i, \mathrm{v}_h, \Phi) \rightarrow \Psi$ *and from* $\vdash \Psi \rightarrow \Phi$ *follows* $\vdash \Psi \rightarrow (\mathrm{Av}_h)S(i, \mathrm{v}_h, \Phi)$.

Proof. From $\vdash \Phi \rightarrow \Psi$ we obtain

$$\vdash S(i, (\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), \Phi \rightarrow \Psi)$$

whence

$$\vdash S(i, (\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), \Phi) \rightarrow S(i, (\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), \Psi).$$

Since a_i does not occur in Ψ, the consequent of the above

implication is equal to Ψ (cf. Chapter II, section 3, theorem f) and hence we obtain $\vdash S(i, (\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), \Phi) \to \Psi$, i.e.

$$\vdash (\mathrm{E}\mathrm{v}_h)S(i, \mathrm{v}_h, \Phi) \to \Psi.$$

If $\vdash \Psi \to \Phi$, then $\vdash \backsim\Phi \to \backsim\Psi$ by the propositional calculus, and hence we obtain

$$\vdash (\mathrm{E}\mathrm{v}_h)S(i, \mathrm{v}_h, \backsim\Phi) \to \backsim\Psi$$

whence $\vdash \Psi \to (\mathrm{A}\mathrm{v}_h)S(i, \mathrm{v}_h, \Phi)$.

Theorems 1 and 2 show that the functional calculus holds in (S). Indeed, in the usual formulation of this calculus one takes the formulas given in the theorem 1 as axioms and rules given in the theorem 2 as rules of inference.

Theorem 3. *If* a_i *occurs and* v_h *does not occur in a matrix* Φ, *then*

$$\vdash (\mathrm{A}\mathrm{v}_h)S(i, \mathrm{v}_h, \backsim\Phi) \to (\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi) \approx 1.$$

This is a restatement of the axiom II 2.

4. Theory of identity. In order to show that this theory holds in (S) it is sufficient to prove the following theorem:

Theorem 1. *If* φ *is a functional form and* Φ *a matrix form and if* $\mathrm{v}_h, \ldots, \mathrm{v}_k$ *are all the bound variables which occur in* φ *or in* Φ *and have there the index* 2, *then*

(1) $$\vdash (\mathrm{A}\mathrm{v}_h) \ldots (\mathrm{A}\mathrm{v}_k)[\mathrm{a}_i \approx \mathrm{a}_j \to \varphi \approx S(i, \mathrm{a}_j, \varphi)],$$

(2) $$\vdash (\mathrm{A}\mathrm{v}_h) \ldots (\mathrm{A}\mathrm{v}_k)\{\mathrm{a}_i \approx \mathrm{a}_j \to [\Phi \leftrightarrow S(i, \mathrm{a}_j, \Phi)]\}.$$

Note that according to theorem II 3 m φ can be represented as

(3) $$\varphi = S^{p}_{\mathrm{v}_h} \ldots S^{q}_{\mathrm{v}_k} \overline{\varphi}$$

and Φ as

(4) $$\Phi = S^{p}_{\mathrm{v}_h} \ldots S^{q}_{\mathrm{v}_k} \overline{\Phi}$$

where $p, \ldots, q$ are arbitrary integers such that the variables $\mathrm{a}_p, \ldots, \mathrm{a}_q$ do not occur in φ or in Φ.

We shall prove theorem 1 using induction on φ and Φ.

If φ is one of the functional forms 1, a_s, v_s $(s = 1, 2, \ldots)$, then the theorem is evident.

Let us assume that φ and ψ are functional forms, Φ and Ψ matrix forms and that $v_s, \ldots, v_t$ are all the bound variables which occur in ψ or in Ψ and have there the index 2. We can represent ψ in the form

$$\psi = S^{x}_{v_s} \ldots S^{z}_{v_t} \overline{\psi} \tag{5}$$

and Ψ in the form

$$\Psi = S^{x}_{v_s} \ldots S^{z}_{v_t} \overline{\Psi} \tag{6}$$

where $x, \ldots, z$ are arbitrary integers such that the variables $a_x, \ldots, a_z$ do not occur in ψ or in Ψ. We may evidently assume that $h = s$ implies $x = p, \ldots,$ and $k = t$ implies $q = z$.

Let us assume that (1), (2), and the following two formulas hold:

$$\vdash (Av_s) \ldots (Av_t)\, [a_i \approx a_j \rightarrow \psi \approx S(i, a_j, \psi)], \tag{7}$$

$$\vdash (Av_s) \ldots (Av_t)\, \{a_i \approx a_j \rightarrow [\Psi \leftrightarrow S(i, a_j, \Psi)]\}. \tag{8}$$

Formulas (1) and (3) imply

$$\vdash a_i \approx a_j \rightarrow \overline{\varphi} \approx S(i, a_j, \overline{\varphi}),$$

and formulas (7) and (5) imply

$$\vdash a_i \approx a_j \rightarrow \overline{\psi} \approx S(i, a_j, \overline{\psi}).$$

Using axioms of the group IV we easily infer that

$$\vdash a_i \approx a_j \rightarrow \overline{\varphi} + \overline{\psi} \approx S(i, a_j, \overline{\varphi} + \overline{\psi})$$

whence it follows by the functional calculus

$$\vdash (Av_h) \ldots (Av_t)[a_i \approx a_j \rightarrow \varphi + \psi \approx S(i, a_j, \varphi + \psi)].$$

In essentially the same way we also prove the formulas

$$\vdash (Av_h) \ldots (Av_t)[a_i \approx a_j \rightarrow \varphi \times \psi \approx S(i, a_j, \varphi \times \psi)],$$

$$\vdash (Av_h) \ldots (Av_t)\{a_i \approx a_j \rightarrow [\varphi \approx \psi \leftrightarrow S(i, a_j, \varphi \approx \psi)]\},$$

$$\vdash (Av_h) \ldots (Av_t)\{a_i \approx a_j \rightarrow [(\Phi \rightarrow \Psi) \leftrightarrow S(i, a_j, \Phi \rightarrow \Psi)]\}.$$

Finally, the formula

$$\vdash (\mathrm{Av}_h) \ldots (\mathrm{Av}_t)[\mathrm{a}_i \approx \mathrm{a}_j \rightarrow (\mu\mathrm{v}_k)\, \Phi \approx S(i, \mathrm{a}_j, (\mu\mathrm{v}_k)\Phi]$$

results from (2) by the functional calculus and the last axiom of the group III.

This proves by induction that the formulas (1) and (2) hold for arbitrary φ in $\mathfrak{F}\mathfrak{f}$, and arbitrary Φ in $\mathfrak{M}\mathfrak{f}$. The proof of theorem 1 is thus complete.

5. Induction. In this section we prove a theorem which states that proofs by induction are valid in (S). We begin by proving the following lemma:

Lemma 1. *If φ is a numerical expression, then*

$$\vdash \backsim \varphi \approx 1 \rightarrow (\varphi \dot{-} 1) + 1 \approx \varphi.$$

Proof. Substituting φ for a in the second axiom of the group IV and changing the bound variable we obtain

$$\vdash \backsim \varphi \approx 1 \rightarrow (\mu\mathrm{v}_h)(\mathrm{v}_h + 1 \approx \varphi) + 1 \approx \varphi$$

whence the lemma follows by definition (8) of section **1**.

Theorem 2. *If Φ is a matrix and*

(1) $$\vdash S(i, 1, \Phi),$$

(2) $$\vdash \Phi \rightarrow S(i, \mathrm{a}_i + 1, \Phi),$$

then $\vdash \Phi$.

Proof. If a_i does not occur in Φ, the theorem is evident, since $S(i, 1, \Phi) = \Phi$. Hence we may assume that a_i occurs in Φ. We denote

$$\varphi = (\mu\mathrm{v}_j)S(i, \mathrm{v}_j, \backsim\Phi)$$

where j is chosen so that neither a_j nor v_j occur in Φ. Using theorem **4** 1 we obtain

(3) $$\vdash \mathrm{a}_i \approx \mathrm{a}_j \rightarrow [\Phi \leftrightarrow S(i, \mathrm{a}_j, \Phi)].$$

Let us perform here the operation of substitution of 1 for the variable a_i. Since $i \neq j$, we have by theorem II **3** h

$$S(i, 1, S(i, \mathrm{a}_j, \Phi)) = S(i, S(i, 1, \mathrm{a}_j), \Phi) = S(i, \mathrm{a}_j, \Phi),$$

and the result of substitution of 1 for a_i in (3) is

$$\vdash 1 \approx a_j \to [S(i, 1, \Phi) \leftrightarrow S(i, a_j, \Phi)].$$

Let us substitute here φ for the variable a_j. Using again theorem II 3 h we obtain easily

$$\vdash 1 \approx \varphi \to [S(i, 1, \Phi) \leftrightarrow S(i, \varphi, \Phi)].$$

Using (1) we obtain now by the propositional calculus

$$(4) \qquad \vdash \backsim S(i, \varphi, \Phi) \to \backsim 1 \approx \varphi.$$

Observe now that by axiom II 3

$$\vdash \varphi \approx (\varphi - 1) + 1 \to \backsim S(i, \varphi - 1, \backsim \Phi)$$

whence by lemma 1

$$\vdash \backsim 1 \approx \varphi \to \backsim S(i, \varphi - 1, \backsim \Phi).$$

Using the law of double negation we obtain further

$$(5) \qquad \vdash \backsim 1 \approx \varphi \to S(i, \varphi - 1, \Phi).$$

Now we substitute $\varphi - 1$ for a_i in (2). Using theorem II 3 h we obtain

$$\vdash S(i, \varphi - 1, \Phi) \to S(i, (\varphi - 1) + 1, \Phi)$$

whence by (5)

$$(6) \qquad \vdash \backsim 1 \approx \varphi \to S(i, (\varphi - 1) + 1, \Phi).$$

To accomplish the proof we substitute in (3) φ for a_i and $(\varphi - 1) + 1$ for a_j. Since a_j occurs neither in φ nor in Φ, we obtain

$$\vdash \varphi \approx (\varphi - 1) + 1 \to [S(i, \varphi, \Phi) \leftrightarrow S(i, (\varphi - 1) + 1, \Phi)].$$

Using (6), lemma 1, and tautologies of the propositional calculus, we get from this formula

$$\vdash \backsim \varphi \approx 1 \to S(i, \varphi, \Phi).$$

Combining this result with (4) we get

$$\vdash \backsim S(i, \varphi, \Phi) \to S(i, \varphi, \Phi)$$

whence (according to the tautology $(\backsim p \to p) \to \backsim \backsim p$)

$$\vdash \backsim \backsim S(i, \varphi, \Phi),$$

i.e. $\vdash \backsim S(i, \varphi, \backsim\Phi)$ since the operations S and $\backsim$ are commutative.

Written without abbreviations this formula becomes

$$\vdash \backsim S(i, (\mu \mathrm{v}_j) S(i, \mathrm{v}_j, \backsim\Phi), \backsim\Phi),$$

i.e. is identical with $\vdash (\mathrm{A}\mathrm{v}_j) S(i, \mathrm{v}_j, \Phi)$. Using theorem 3 1 we obtain finally $\vdash \Phi$.

The proof of theorem 2 is thus complete.

Theorem 2 shows that the usual form of the inductive proofs is translatable into the system (S). We can therefore prove in (S) the usual arithmetical theorems imitating the proofs which are given e.g. in the textbook of Landau [14]. We do not need to enter deeper into details of this reconstruction of intuitive arithmetic in the system (S) and shall content ourselves with some examples which we shall discuss in the next sections.

6. Addition and multiplication of digits. In this section we prove two seemingly obvious but very important theorems.

Theorem 1. $\vdash D_n + D_m \approx D_{n+m}$, $\vdash D_n \times D_m \approx D_{nm}$.

Proof. We use induction on m. If $m = 1$, then $D_n + D_1 = D_n + 1 = D_{n+1}$ according to the definition of D_n given in II 3. Using axiom III 1 we obtain therefore $\vdash D_n + D_1 \approx D_{n+1}$.

Let us assume that for an m

$$\vdash D_n + D_m \approx D_{n+m}.$$

Since $D_{m+1} = D_m + 1$, we obtain

$$D_n + D_{m+1} = D_n + (D_m + 1)$$

and hence we obtain by axiom IV 2

$$\vdash D_n + D_{m+1} \approx (D_n + D_m) + 1.$$

Using the inductive hypothesis we obtain from this formula

$$\vdash D_n + D_{m+1} \approx D_{n+m} + 1$$

and hence, according to the definition of the function D,

$$\vdash D_n + D_{m+1} \approx D_{n+m+1}.$$

This completes the proof of the first formula given in the theorem. The proof of the second formula is similar and can be omitted here.

Theorem 2. $\vdash \backsim \mathrm{a} \approx \mathrm{a} + \mathrm{b}$.

Proof. Put $\Phi = \backsim \mathrm{a} \approx \mathrm{a} + \mathrm{b}$. By axiom IV 1 we have $\vdash S(1, 1, \Phi)$. Since

$$\vdash \mathrm{a} + 1 \approx (\mathrm{a} + 1) + \mathrm{b} \rightarrow \mathrm{a} \approx \mathrm{a} + \mathrm{b},$$

as we easily see from the axioms IV 2, IV 3, and IV 4, we obtain by the propositional calculus

$$\vdash \Phi \rightarrow S(1, \mathrm{a} + 1, \Phi).$$

Using theorem 5 2 we obtain now $\vdash \Phi$, i.e., $\vdash \backsim \mathrm{a} \approx \mathrm{a} + \mathrm{b}$, q.e.d.

Corollary 3. *If* $n \neq m$, *then* $\vdash \backsim D_n \approx D_m$.

Proof. We can evidently assume that $n < m$. Substitute D_n and D_{m-n} for a and b in theorem 2. Using theorem 1 we obtain then at once the desired result.

7. Theorems on inequalities.

Theorem 1. $\vdash \mathrm{a} \lessdot \mathrm{a} + 1$.

Proof. Put $\Phi = \mathrm{a} + \mathrm{b} \approx \mathrm{a} + 1$. By axiom II 1 we obtain

$$\vdash \Phi \rightarrow \mathrm{a} + (\mu \mathrm{x}) S(2, \mathrm{x}, \Phi) \approx \mathrm{a} + 1$$

whence $\vdash \mathrm{a} + \mathrm{b} \approx \mathrm{a} + 1 \rightarrow \mathrm{a} \lessdot \mathrm{a} + 1$. Substituting here 1 for the variable b and applying axiom III 1 we obtain the desired result by the rule of modus ponens.

Theorem 2. $\vdash \mathrm{a} \lessdot \mathrm{b} \rightarrow (\mathrm{b} \lessdot \mathrm{c} \rightarrow \mathrm{a} \lessdot \mathrm{c})$.

Proof. Using theorem 4 1 we obtain easily

$$\vdash \mathrm{a} + \mathrm{d} \approx \mathrm{b} \rightarrow [\mathrm{b} + \mathrm{e} \approx \mathrm{c} \rightarrow \mathrm{a} + (\mathrm{d} + \mathrm{e}) \approx \mathrm{c}].$$

By axiom II 1

$$\vdash \mathrm{a} + (\mathrm{d} + \mathrm{e}) \approx \mathrm{c} \rightarrow \mathrm{a} + (\mu \mathrm{x})(\mathrm{a} + \mathrm{x} \approx \mathrm{c}) \approx \mathrm{c}$$

whence

$$\vdash a + d \approx b \rightarrow (b + e \approx c \rightarrow a \lessdot c).$$

Substituting here $(\mu x)(a + x \approx b)$ for d and $(\mu x)(b + x \approx c)$ for e we obtain the theorem.

Theorem 3. $\vdash a \lessdot a + b$.

Proof. Put $\Phi = a \lessdot a + b$. Theorem 1 can be written as $\vdash S(2, 1, \Phi)$, and by theorems 1 and 2 we easily obtain

$$\vdash \Phi \rightarrow S(2, b + 1, \Phi).$$

Using theorem 5 2 we obtain therefore $\vdash \Phi$ which was to be proved

Theorem 4. *If* $n < m$ *then* $\vdash D_n \lessdot D_m$.

Proof. By theorem 3 $\vdash D_n \lessdot D_n + D_{m-n}$. Since $\vdash D_m \approx D_n + D_{m-n}$ we obtain the theorem using the general theorem on identity established in section 4.

Theorem 5. $\vdash a + 1 \lessdot b + 1 \rightarrow a \lessdot b$.

Proof. By the axioms of group IV we have

$$\vdash (a + 1) + c \approx b + 1 \rightarrow a + c \approx b$$

whence

$$\vdash (a + 1) + c \approx b + 1 \rightarrow a + (\mu x)(a + x \approx b) \approx b.$$

Substituting here $(\mu x)[(a + 1) + x \approx b + 1]$ for c we obtain the desired result.

Theorem 6. $\vdash (a + b \approx c + 1) \rightarrow [a \lessdot c \vee a \approx c]$.

Proof. In virtue of theorem 4 1 we obtain

$$(1) \qquad \vdash b \approx 1 \rightarrow (a + b \approx c + 1 \rightarrow a \approx c).$$

Using again theorem 4 1 we obtain

$$\vdash b \approx (b - 1) + 1 \rightarrow [a + b \approx c + 1 \rightarrow a + (b - 1) \approx c]$$

and hence

$$(2) \qquad \vdash b \approx (b - 1) + 1 \rightarrow [a + b \approx c + 1 \rightarrow a \lessdot c]$$

since $\vdash \mathrm{a} \prec \mathrm{a} + (\mathrm{b} - 1)$ by theorem 3. From (1) and (2) we obtain by the propositional calculus

$$\vdash [\backsim \mathrm{b} \approx 1 \rightarrow \mathrm{b} \approx (\mathrm{b} - 1) + 1] \rightarrow [\mathrm{a} + \mathrm{b} \approx \mathrm{c} + 1 \rightarrow \rightarrow (\mathrm{a} \prec \mathrm{c} \vee \mathrm{a} \approx \mathrm{c})]$$

whence the theorem follows in virtue of lemma **5** 1.

Theorem 7. $\vdash \mathrm{a} \prec D_{n+1} \leftrightarrow (\mathrm{a} \approx D_1 \vee \mathrm{a} \approx D_2 \vee \ldots \vee \mathrm{a} \approx D_n)$.

Proof. The implication from right to left results immediately from theorem 4. The converse implication will be proved by induction on n.

Let us first assume that $n = 1$. From theorems 6, **6** 3, and the remark that $\vdash \backsim \mathrm{a} \prec D_1$ we obtain

$$\vdash \mathrm{a} + \mathrm{c} \approx D_2 \rightarrow \mathrm{a} \approx D_1$$

and hence by substitution $\vdash \mathrm{a} \prec D_2 \rightarrow \mathrm{a} \approx D_1$. Theorem 7 is thus proved for $n = 1$.

Let us now assume that the theorem is true for $n = k - 1$. Since $\vdash \mathrm{a} + \mathrm{c} \approx D_{k+1} \rightarrow (\mathrm{a} \prec D_k \vee \mathrm{a} \approx D_k)$ we obtain in virtue of the inductive hypothesis

$$\vdash \mathrm{a} + \mathrm{c} \approx D_{k+1} \rightarrow (\mathrm{a} \approx D_1 \vee \mathrm{a} \approx D_2 \vee \ldots \vee \mathrm{a} \approx D_k).$$

whence the theorem follows by substitution of $(\mu\mathrm{x})(\mathrm{a} + \mathrm{x} \approx D_{k+1})$ for the variable c.

Theorem 8. *If* $m \leqslant n$, *then* $\vdash \backsim D_n \prec D_m$.

Proof. According to the corollary **6** 3 $\vdash \backsim D_n \approx D_i$ for $i = 1, 2, \ldots, m - 1$ whence we obtain

$$\vdash \backsim [D_n \approx D_1 \vee \ldots \vee D_n \approx D_{m-1}].$$

This gives us $\vdash \backsim D_n \prec D_m$ in virtue of theorem 7 and known tautologies of the propositional calculus.

Theorem 9. $\vdash 1 \prec \mathrm{a} \vee 1 \approx \mathrm{a}$.

Proof. It follows from lemma **5** 1 that

$$\vdash \mathrm{a} \approx 1 \vee (\mathrm{a} - 1) + 1 \approx \mathrm{a}$$

whence the desired result follows by means of theorem 3 and axioms of the group IV.

Theorem 10. $\vdash b \lessdot a \rightarrow [b + 1 \lessdot a \vee b + 1 \approx a]$.

Proof. It is easy to show that $\vdash b \lessdot a \rightarrow \backsim a \approx 1$ which gives in virtue of lemma **5** 1

$$(3) \qquad \vdash b \lessdot a \rightarrow (a - 1) + 1 \approx a.$$

If we denote $(\mu x)(b + x \approx a)$ by φ we can write this formula in the form

$$\vdash b + \varphi \approx a \rightarrow b + \varphi \approx (a - 1) + 1.$$

Using theorem 6 we obtain now

$$\vdash b \lessdot a \rightarrow [b \lessdot a - 1 \vee b \approx a - 1].$$

Since $\vdash b \lessdot a - 1 \rightarrow b + 1 \lessdot (a-1) + 1$ and $\vdash b \approx a - 1 \rightarrow b + 1 \approx (a - 1) + 1$, we obtain

$$\vdash b \lessdot a \rightarrow [b + 1 \lessdot (a - 1) + 1 \vee b + 1 \approx (a - 1) + 1]$$

and the theorem follows in virtue of (3).

Theorem 11. $\vdash b \lessdot a \vee b \approx a \vee a \lessdot b$.

Proof. Put $\Phi = b \lessdot a \vee b \approx a \vee a \lessdot b$. Theorem 9 can be written in the form $\vdash S(2, 1, \Phi)$. Since $\vdash b \approx a \vee a \lessdot b \rightarrow a \lessdot b + 1$, we obtain

$$\vdash b \approx a \vee a \lessdot b \rightarrow S(2, b + 1, \Phi).$$

By theorem 10 we have further $\vdash b \lessdot a \rightarrow S(2, b + 1, \Phi)$. Combining this with the previous formula we obtain by the propositional calculus $\vdash \Phi \rightarrow S(2, b + 1, \Phi)$, and hence theorem **5** 2 yields $\vdash \Phi$, q.e.d.

Theorem 12. *If Φ is a matrix in which exactly one free variable a_i occurs and if v_h is a bound variable which does not occur in Φ, then*

$$\vdash S(i, D_q, \Phi) \,\&\, S(i, D_1, \backsim\Phi) \,\&\, \ldots \,\&\, S(i, D_{q-1}, \backsim\Phi) \rightarrow (\mu v_h) S(i, v_h, \Phi) \approx D_q.$$

Proof. Let us put $\varphi = (\mu v_h)S(i, v_h, \Phi)$ and denote by H the antecedent of the formula to be proved. From axiom II 1 we obtain

(4) $$\vdash S(i, D_q, \Phi) \to S(i, \varphi, \Phi).$$

Since $\vdash \varphi \approx D_j \to [S(i, \varphi, \Phi) \leftrightarrow S(i, D_j, \Phi)]$ by theorem 4 1, we obtain using tautologies of the propositional calculus

$$\vdash S(i, D_1, \backsim\Phi) \;\&\; \ldots \;\&\; S(i, D_{q-1}, \backsim\Phi) \;\&\; S(i, \varphi, \Phi) \\ \backsim(\varphi \approx D_1) \;\&\; \ldots \;\&\; \backsim(\varphi \approx D_{q-1})$$

whence on account of (4) and theorem 7

(5) $$\vdash H \to \backsim(\varphi \prec D_q).$$

From axiom II 3 we obtain further

$$\vdash (\varphi \approx D_q + a_j) \to \backsim S(i, D_q, \Phi)$$

whence by contraposition $\vdash H \to \backsim(\varphi \approx D_q + a_j)$. If we substitute here $(\mu v_h)(D_q + v_h \approx \varphi)$ for the variable a_j, we obtain

(6) $$\vdash H \to \backsim(D_q \prec \varphi).$$

Theorem 12 results now from (5), (6), and theorem 11.

8. Matrices satisfied by exactly one digit. We shall prove in this section the following theorem:

Theorem 1. *Let Φ and Ψ be two matrices in which exactly one free variable a_1 occurs and let v_k be a bound variable which occurs neither in Φ nor in Ψ. Assume that*

(1) $$\vdash \Phi \;\&\; S(1, a_2, \Phi) \to a_1 \approx a_2,$$

(2) $$\vdash S(1, D_n, \Phi).$$

Under these assumptions

(3) $$\vdash (\mathrm{A}v_k)S(1, v_k, \Phi \to \Psi) \leftrightarrow S(1, D_n, \Psi).$$

Proof. From theorem 3 1 we obtain

$$\vdash (\mathrm{A}v_k)S(1, v_k, \Phi \to \Psi) \to [S(1, D_n, \Phi) \to S(1, D_n, \Psi)]$$

and hence by (2)

$$(4)\qquad \vdash (\mathrm{Av}_k)S(1, \mathrm{v}_k, \Phi \to \Psi) \to S(1, D_n, \Psi).$$

To prove the converse implication we proceed as follows. From (2) we obtain

$$\vdash \Phi \to \Phi \;\&\; S(1, D_n, \Phi)$$

and hence on account of (1) $\vdash \Phi \to \mathrm{a}_1 \approx D_n$. Multiplying both sides of this implication by $S(1, D_n, \Psi)$ we obtain

$$(5)\qquad \vdash \Phi \;\&\; S(1, D_n, \Psi) \to (\mathrm{a}_1 \approx D_n) \;\&\; S(1, D_n, \Psi)$$

Theorem 4 1 yields

$$\vdash \mathrm{a}_1 \approx D_n \;\&\; S(1, D_n, \Psi) \to \Psi$$

and hence we obtain from (5) $\vdash \Phi \;\&\; S(1, D_n, \Psi) \to \Psi$, or what is essentially the same $\vdash S(1, D_n, \Psi) \to (\Phi \to \Psi)$. Applying now theorem 3 2 we obtain

$$(6)\qquad \vdash S(1, D_n, \Psi) \to (\mathrm{Av}_k)S(1, \mathrm{v}_k, \Phi \to \Psi).$$

Theorem 1 results now immediately from (4) and (6).

SEMANTICS OF (S) [1]

1. Representability of infinite sequences. In the present Chapter we shall work with infinite sequences almost all [2] terms of which are equal to 1. Every such sequence $a_1, a_2, \ldots, a_n, 1, 1, \ldots$ can conveniently be represented by the integer

$$m = p_1^{a_1-1} \cdot p_2^{a_2-1} \ldots p_n^{a_n-1}$$

where p_i is the i-th prime. Conversely, every integer m determines a sequence almost all terms of which are equal to 1. In order to obtain this sequence we put

$$a_n = W(p_n, m)$$

where $W(x, y)$ is the least integer z such that y is not divisible by x^z. We shall also use the shorter symbol $\overline{\overline{m}}_n$ instead of $W(p_n, m)$.

We shall often use the auxiliary function

$$C_i(m, a) = m \cdot p_i^{a-\overline{\overline{m}}_i}.$$

It follows from this definition that

$$W(p_n, C_i(m, a)) = W(p_n, m) \text{ if } i \neq n$$
$$W(p_i, C_i(m, a)) = a.$$

The function C enables us therefore to construct from a given integer m another integer representing a sequence which differs but in the i-th term from the sequence represented by m.

Note the following useful property of the function C:

(1) *If* $i \neq j$, *then* $C_i(C_j(m, a), b) = C_j(C_i(m, b), a)$.

[1] This Chapter is based entirely on works of Tarski. Cf. his papers [21] and [22].

[2] I.e. all with an exception of at most finite number.

2. Values of functional forms and the notion of satisfaction for matrix forms. Before we give exact definitions of these basic semantical notions we shall explain briefly their intuitive content. Let φ be a functional form, e.g.

$$\varphi = \varphi' = \mathrm{v}_1 + (\mathrm{a}_1 \times \mathrm{a}_2) \text{ or } \varphi = \varphi'' = (\mu \mathrm{v}_1)(\mathrm{v}_1 \times \mathrm{v}_1 \approx \mathrm{a}_1).$$

Speaking intuitively, every such form represents an arithmetical function of as many variables as there are free variables and bound variables with the index 2 occurring in φ. For instance φ' represents the function $F(x, y, z) = x + y \cdot z$ and φ'' represents the function $F(x)$ whose value is 1 if x is not a square of an integer and which is equal to y if $x = y^2$.

Let us ascribe arbitrary numerical values to free variables occurring in φ and also to the bound variables which occur in φ and have therein the index 2. The function F represented by φ takes then on a numerical value which we shall call the *value* of φ for the given values of the variables. For instance, the value of φ' is 7 if we ascribe the value 3 to v_1 and the values 2 to a_1 and a_2. If we ascribe the value 2 to a_1, then the value of φ'' is 1; if the value of a_1 is 4, then the value of φ'' is 2.

Let now Φ be a matrix form, e.g. $\varphi' \approx \varphi'' \rightarrow \varphi' \approx 1$ and let us ascribe arbitrary values to the free variables occurring in Φ as well as to the bound variables which occur in Φ and have therein the index 2. The intuitive meaning of Φ is this: If the values of φ' and φ'' are equal, then the value of φ' is 1. Hence Φ represents a theorem of arithmetic and this theorem can be either true or false. In the first case we say that the values given to the variables *satisfy* the matrix form Φ, and in the second that these values *do not satisfy* Φ. For instance in the example considered above the values 2, 1, 2 given to the variables v_1, a_1, a_2 satisfy Φ.

It is convenient to ascribe values to all variables simultaneously, independently of whether they occur in the expression which we consider or not. Every such system of values can be identified with an infinite sequence $a_1, a_2, \ldots$ in which a_{2n-1} is the value given to the variable v_n and a_{2n} the value given to the variable a_n. Since we shall never deal simultaneously with an

infinite number of expressions, we do not need to consider wholly arbitrary sequences but can limit ourselves to sequences with almost all terms equal to 1. Every such sequence can be represented by an integer in the way explained in section 1. It follows that the value of a functional form φ is a function of φ and of an integer m which synthetizes the values ascribed to the variables. For the same reason the notion of satisfaction is a binary relation between matrix-forms and integers.

We shall denote by $Val(\varphi, m)$ the value of φ for values of variables represented by the integer m and shall write $\mathfrak{Stsf}(\Phi, m)$ instead of "*the values of variables represented by* m *satisfy* Φ".

3. Inductive definition of $Val\,(\varphi, m)$ **and** $\mathfrak{Stsf}\,(\Phi, m)$. In this section we shall give an exact definition of the notions which were explained intuitively in the previous section. To obtain this definition we first define $Val(\varphi, m)$ for the simplest functional forms a_h, v_h, and 1 and then define it for the functional forms $\varphi \boldsymbol{+} \psi$, $\varphi \times \psi$ under the assumption that it has been defined for the functional forms φ and ψ. Under the same assumption we define also the meaning of the formula $\mathfrak{Stsf}(\varphi \approx \psi, m)$. Next we define the meaning of the formula $\mathfrak{Stsf}(\Phi \rightarrow \Psi, m)$ under the assumption that the meaning of the formulas $\mathfrak{Stsf}(\Phi, m)$ and $\mathfrak{Stsf}(\Psi, m)$ are already defined. Finally we define $Val((\mu \mathrm{v}_h)\Phi, m)$ under the assumption that the meaning of the formula $\mathfrak{Stsf}(\Phi, m')$ is already defined for an arbitrary integer m'. In this way $Val(\varphi, m)$ and $\mathfrak{Stsf}(\Phi, m)$ will be defined for arbitrary φ in $\mathfrak{Ff}$ and Φ in $\mathfrak{Mf}$.

We divide our inductive definition into eight parts:

(1) $Val(1, m) = 1$,

(2) $Val(\mathrm{v}_h, m) = \overline{\overline{m}}_{2h-1}$,

(3) $Val(\mathrm{a}_h, m) = \overline{\overline{m}}_{2h}$,

(4) $Val(\varphi \boldsymbol{+} \psi, m) = Val(\varphi, m) + Val(\psi, m)$,

(5) $Val(\varphi \times \psi, m) = Val(\varphi, m) \cdot Val(\psi, m)$,

(6) $\mathfrak{Stsf}(\varphi \approx \psi, m) \equiv Val(\varphi, m) = Val(\psi, m)$,

(7) $\mathfrak{Stsf}(\Phi \rightarrow \Psi, m) \equiv \mathfrak{Stsf}(\Phi, m) \supset \mathfrak{Stsf}(\Psi, m)$,

(8) $Val((\mu \mathrm{v}_h)\Phi, m) = 1$ *if there is no a such that* $\mathfrak{Stsf}(\Phi, C_{2h-1}(m, a))$, otherwise

$Val((\mu \mathrm{v}_h)\Phi, m) =$ *the least such integer a.*

To illustrate how this definition works we take

$$\varphi = (\mu \mathrm{v}_1)[(\mathrm{v}_1 \times \mathrm{v}_1) \approx ((\mathrm{a}_1 \times \mathrm{a}_1) \mathbin{\text{+}} 1)]$$

and calculate $Val(\varphi, m)$. By (2), (3), (4), and (5) we obtain $Val(\mathrm{v}_1 \times \mathrm{v}_1, m) = \overline{\overline{m}}_1^2$, $Val((\mathrm{a}_1 \times \mathrm{a}_1) \mathbin{\text{+}} 1, m) = \overline{\overline{m}}_2^2 + 1$ whence by (6)

$$\mathfrak{Stsf}(((\mathrm{v}_1 \times \mathrm{v}_1) \approx (\mathrm{a}_1 \times \mathrm{a}_1) \mathbin{\text{+}} 1),\ C_1(m, a)) \equiv (a^2 = \overline{\overline{m}}_2^2 + 1).$$

Since $x^2 + 1$ is never a square, we infer that there is no a such that

$$\mathfrak{Stsf}(((\mathrm{v}_1 \times \mathrm{v}_1) \approx (\mathrm{a}_1 \times \mathrm{a}_1) \mathbin{\text{+}} 1),\ C_1(m, a))$$

and consequently $Val(\varphi, m) = 1$ according to (8).

One should not be deceived by the superficial similarity of the inductive definition given in this section and the inductive definition given in Chapter II, section 2, p. 28. To explain the chief difference between these definitions we remark the following.

In both definitions we have integers on which the induction proceeds (they are denoted by "b" and "c" in Chapter II, and by "φ" and "Φ" in the present Chapter). Furthermore, we have in both definitions the parameters ("a", "i", "p" in Chapter II, and "m" in the present Chapter). Now the parameters in the definitions of Chapter II are kept constant whereas in the present definition they are variable (cf. (8)). This has the effect that we can calculate the values of functions defined in Chapter II by tracing backwards the steps of the definition, and arriving to an end after a finite number of steps. E.g. if we have to calculate $Oc(a, b)$ we try to decompose b into simpler constituents; if we find for instance that $b = b_1 \to b_2$, we reduce our problem to a calculation of $Oc(a, b_1)$ and $Oc(a, b_2)$. After a finite number of such steps we arrive finally to the values $Oc(a, 1)$, $Oc(a, \mathrm{v}_h)$, and $Oc(a, \mathrm{a}_h)$ which are given explicitly and obtain then the value of $Oc(a, b)$ by repeated substitutions.

The situation is entirely different in case of the definition given

in the present Chapter. Indeed, if $\varphi = (\mu v_h)\Phi$, then the calculation of $Val(\varphi, m)$ is reduced to that of $\mathfrak{Stg}(\Phi, m')$ for infinitely many different values of m'. Hence the calculation of $Val(\varphi, m)$ cannot be completed in a finite number of steps and the definition given in the present Chapter has an "infinitary" character which distinguishes it essentially from the superficially similar "finitary" definitions of Chapter II.

Because of the infinitary character of the definition (1) — (8) we cannot expect that the same method which we used in Chapter II will allow us to replace the inductive definition by an explicit one. As a matter of fact it can be shown that an explicit definition of Val and $\mathfrak{Stg}$ is possible only when we use the general notion of an arbitrary set of integers.

In the next section we shall outline an explicit definition of Val and $\mathfrak{Stg}$. The proof that it cannot be replaced by a purely arithmetical one (i.e. such which avoids the notion of an arbitrary set) will be given in Chapter VI.

4. Explicit definition of *Val* and $\mathfrak{Stg}$ [3]. We shall say that a set $\mathfrak{X}$ of integers *represents a function* if

(1) *for every n there is in $\mathfrak{X}$ an x such that $K_1(x) = n$,*
(2) *for every x and y in $\mathfrak{X}$ if $K_1(x) = K_1(y)$, then $K_2(x) = K_2(y)$.*

To explain this definition we remark that a function F can be identified with the set of ordered pairs $(n, F(n))$ and a pair (a, b) can be identified with the integer $J(a, b)$ (cf. Chapter I, p. 14). In this way a function F is converted into a set $\mathfrak{X}$ and it is easy to show that this set must satisfy the conditions (1) and (2).

For every Φ in $\mathfrak{Mk}$ let us denote by $\mathfrak{A}(\Phi)$ the set of those m for which $\mathfrak{Stg}(\Phi, m)$. Similarly for every φ in $\mathfrak{Fk}$ let us denote by $\mathfrak{A}(\varphi)$ the set representing the function $Val(\varphi, m)$ treated as the function of m alone.

It can be easily shown that $\mathfrak{A}(1)$ is the set of all integers having the form $J(m, 1)$:

$$x \in \mathfrak{A}(1) \equiv K_2(x) = 1.$$

[3] Cf. Tarski [20], pp. 311—312.

It is also easy to prove that $\mathfrak{A}(\mathrm{v}_h)$ consists of all integers of the form $J(m, \overline{\overline{m}}_{2h-1})$ i.e. of the form $J(m, W(p_{2h-1}, m))$, and that $\mathfrak{A}(\mathrm{a}_h)$ consists of all integers of the form $J(m, \overline{\overline{m}}_{2h}) = J(m, W(p_{2h}, m))$:

$$x \in \mathfrak{A}(\mathrm{v}_h) \equiv K_2(x) = W(p_{2h-1}, K_1(x)),$$
$$x \in \mathfrak{A}(\mathrm{a}_h) \equiv K_2(x) = W(p_{2h}, K_1(x)).$$

We introduce further certain operations on sets of integers which we denote by symbols similar to those used for operations on expressions:

$$\mathfrak{B} \dot{+} \mathfrak{C} = \lambda x(\exists b)(\exists c)[(b \in \mathfrak{B})\cdot(c \in \mathfrak{C})\cdot$$
$$(K_1(b) = K_1(c) = K_1(x))\cdot(K_2(b) + K_2(c) = K_2(x))],$$
$$\mathfrak{B} \dot{\times} \mathfrak{C} = \lambda x(\exists b)(\exists c)[(b \in \mathfrak{B})\cdot(c \in \mathfrak{C})\cdot$$
$$(K_1(b) = K_1(c) = K_1(x))\cdot(K_2(b)\cdot K_2(c) = K_2(x))],$$
$$\mathfrak{B} \dot{\approx} \mathfrak{C} = \lambda x(\exists y)(J(x, y) \in \mathfrak{B} \wedge \mathfrak{C}),$$
$$\mathfrak{B} \dot{\rightarrow} \mathfrak{C} = -\mathfrak{B} \vee \mathfrak{C},$$
$$\mathfrak{M}_h\mathfrak{B} = \lambda x[K_2(x) = \min_a(C_{2h-1}\,(K_1(x), a) \in \mathfrak{B} \textit{ or } K_2(x) = 1$$
$$\textit{provided that no such } a \textit{ exists}].$$

It can be shown without essential difficulties that

$$\mathfrak{A}(\varphi + \psi) = \mathfrak{A}(\varphi) \dot{+} \mathfrak{A}(\psi),$$
$$\mathfrak{A}(\varphi \times \psi) = \mathfrak{A}(\varphi) \dot{\times} \mathfrak{A}(\psi),$$
$$\mathfrak{A}(\varphi \approx \psi) = \mathfrak{A}(\varphi) \dot{\approx} \mathfrak{A}(\psi),$$
$$\mathfrak{A}(\Phi \rightarrow \Psi) = \mathfrak{A}(\Phi) \dot{\rightarrow} \mathfrak{A}(\Psi),$$
$$\mathfrak{A}((\mu \mathrm{v}_h)\Phi) = \mathfrak{M}_h\mathfrak{A}(\Phi).$$

Let now Ω be an arbitrary expression. It follows easily from the definition of the class of expressions that there exists a sequence of expressions

$$\text{(i)}\quad \Omega_1, \Omega_2, \ldots, \Omega_n = \Omega$$

and a sequence

$$\text{(ii)}\quad i_1, i_2, \ldots, i_n$$

whose elements are integers 1 or 2 such that for every $j \leqslant n$ one of the following conditions is satisfied:

(iii) $\Omega_j = 1$ or $\Omega_j \in \mathfrak{Vb}$ or $\Omega_j \in \mathfrak{Vf}$ and $i_j = 1$,

(iv) $\Omega_j = \Omega_k + \Omega_l$ or $\Omega_j = \Omega_k \times \Omega_l$ and $i_j = i_k = i_l = 1 \quad (k < j,\ l < j)$,

(v) $\Omega_j = \Omega_k \approx \Omega_l$ and $i_j = 2,\ i_k = i_l = 1 \quad (k < j,\ l < j)$,

(vi) $\Omega_j = \Omega_k \rightarrow \Omega_l$ and $i_j = i_k = i_l = 2 \quad (k < j,\ l < j)$,

(vii) $\Omega_j = (\mu \mathrm{v}_h)\Omega_k$ and $i_j = 1,\ i_k = 2 \quad (j < k)$.

Put $\mathfrak{A}_j = \mathfrak{A}(\Omega_j)$ for $j = 1, 2, \ldots, n$. We obtain thus a sequence of sets such that

(viii) In the cases (iii) $\mathfrak{A}_j$ is either $\mathfrak{A}(1)$, or $\mathfrak{A}(\mathrm{v}_h)$, or $\mathfrak{A}(\mathrm{a}_h)$.

(ix) In the cases (iv) $\mathfrak{A}_j$ is either $\mathfrak{A}_k \dot{+} \mathfrak{A}_l$ or $\mathfrak{A}_k \dot{\times} \mathfrak{A}_l$.

(x) In the case (v) $\mathfrak{A}_j$ is $\mathfrak{A}_k \dot{\approx} \mathfrak{A}_l$.

(xi) In the case (vi) $\mathfrak{A}_j$ is $\mathfrak{A}_k \dot{\rightarrow} \mathfrak{A}_l$.

(xii) In the case (vii) $\mathfrak{A}_j$ is $\mathfrak{M}_h\, \mathfrak{A}_k$.

Conversely, if a sequence

$$\text{(xiii)} \quad \mathfrak{A}_1, \mathfrak{A}_2, \ldots, \mathfrak{A}_n$$

of sets satisfies conditions (viii) — (xii), then $\mathfrak{A}_j = \mathfrak{A}(\Omega_j)$ for $j = 1, 2, \ldots, n$.

Observing that if Ω is a matrix form, then

$$\mathfrak{Stf}(\Omega, m) \equiv m \in \mathfrak{A}(\Omega),$$

and if Ω is a functional form, then

$$Val(\Omega, m) = \textit{the } x \textit{ for which } J(m, x) \in \mathfrak{A}(\Omega)$$

we can express the explicit definition of *Val* and $\mathfrak{Stf}$ as follows:

> $x = Val(\Omega, m)$ *if and only if there exist sequences* (i), (ii), (xiii) *satisfying conditions* (iii) — (xii) *and such that* $i_n = 1$ *and* $J(m, x) \in \mathfrak{A}_n$;
>
> $\mathfrak{Stf}(\Omega, m)$ *if and only if there exist sequences* (i), (ii), (xiii) *satisfying conditions* (iii) — (xii) *and such that* $i_n = 2$ *and* $m \in \mathfrak{A}_n$.

The logical form of these definitions could be simplified by an identification of sequences of integers with integers representing these sequences (cf. Chapter I, section 2, p. 15). Furthermore, we can identify finite sequences of sets of integers with single sets: Instead of the sequence (xiii) we can consider the set $\mathfrak{X}$ of integers g with the properties

$$L(g) = n, \ \bar{g}_i \in \mathfrak{A}_i \text{ for } i = 1, 2, \ldots, L(g).$$

We omit the details of these simplifications since they are very easy and not essential for our further purpose.

5. Class $\mathfrak{Tr}$. A matrix form Φ will be said to belong to the class $\mathfrak{Tr}$ if every integer satisfies Φ:

$$\Phi \in \mathfrak{Tr} \equiv (\Phi \in \mathfrak{Mf}) \cdot (m) \mathfrak{Stsf}(\Phi, m).$$

In case when Φ is a sentence, we shall often say "*Φ is true*" instead of "*Φ is in the class $\mathfrak{Tr}$*". We shall abstain however from using the word "true" in cases when Φ contains free variables. Instead of $\Phi \in \mathfrak{Tr}$ we shall write sometimes $\models\!\!\!\vdash \Phi$.

An example of a matrix which belongs to the class $\mathfrak{Tr}$ is $\Phi = \varphi \approx 1$ where

$$\varphi = (\mu \mathrm{v}_1)((\mathrm{v}_1 \times \mathrm{v}_1) \approx (\mathrm{a}_1 \times \mathrm{a}_1) + 1).$$

Indeed, we have seen in section 3, p. 59 that $Val(\varphi, m) = 1$ for every m, and since $Val(1, m) = 1$, we obtain $Val(\varphi, m) = Val(1, m)$ and hence $\mathfrak{Stsf}(\Phi, m)$ for every m.

Theorem 1. *If Φ and Ψ are in $\mathfrak{Tr}$, then so is $Mp(\Phi, \Psi)$.*

Proof. We can assume that $\Phi = \Psi \to \Omega$ since otherwise the theorem is evident. Using the formula (7) of section 3 and the assumptions that

$$\mathfrak{Stsf}(\Psi, m) \text{ and } \mathfrak{Stsf}(\Psi \to \Omega, m)$$

we obtain $\mathfrak{Stsf}(\Omega, m)$ whence $\Omega \in \mathfrak{Tr}$. Since $\Omega = Mp(\Phi, \Psi)$ the theorem is proved.

Theorem 1 says that the class $\mathfrak{Tr}$ is closed under the rule (or

better the operation) of "modus ponens". We shall now prove the same for the operation of substitution. To achieve this we need two lemmas:

Lemma 2. *If Ω is an expression and m', m'' two integers satisfying the following conditions*

$$\text{if } \mathrm{a}_j \text{ occurs in } \Omega, \text{ then } \overline{\overline{m}}'_{2j} = \overline{\overline{m}}''_{2j},$$
$$\text{if } Ind(\mathrm{v}_j, \Omega) = 2, \text{ then } \overline{\overline{m}}'_{2j-1} = \overline{\overline{m}}''_{2j-1},$$

then

(1) $\mathfrak{Stsf}(\Omega, m') \equiv \mathfrak{Stsf}(\Omega, m'')$ *if* Ω *is in* $\mathfrak{Mf}$,

(2) $Val(\Omega, m') = Val(\Omega, m'')$ *if* Ω *is in* $\mathfrak{Ff}$.

Proof. We use induction on Ω. If $\Omega = 1$, equation (2) is evident. If $\Omega = \mathrm{v}_h$, then $\overline{\overline{m}}'_{2h-1} = \overline{\overline{m}}''_{2h-1}$ since $Ind(\mathrm{v}_h, \mathrm{v}_h) = 2$, and hence the equation (2) is satisfied because its left hand side is equal to $\overline{\overline{m}}'_{2h-1}$ and its right hand side to $\overline{\overline{m}}''_{2h-1}$. The proof in the case $\Omega = \mathrm{a}_h$ is similar.

Assume now that (2) holds for two functional forms φ and ψ and let Ω be one of the expressions $\varphi + \psi$, $\varphi \times \psi$, $\varphi \approx \psi$. If m' and m'' satisfy the assumptions of the lemma with respect to Ω they do so with respect to φ and ψ since $Oc(a, \varphi) = 2$ or $Oc(a, \psi) = 2$ implies that $Oc(a, \Omega) = 2$ and $Ind(\mathrm{v}_j, \varphi) = 2$ or $Ind(\mathrm{v}_j, \psi) = 2$ implies that $Ind(a, \Omega) = 2$. Hence the equation (2) holds for the functional forms φ and ψ, and we easily infer that this equation holds also for the functional forms $\varphi + \psi$, $\varphi \times \psi$ and for the matrix form $\varphi \approx \psi$.

In a similar way we show that if (1) holds for two matrix forms Φ and Ψ, it does so for the matrix form $\Phi \rightarrow \Psi$.

Let us finally assume that (1) holds for the matrix form Φ, and let Ω be the functional form $(\mu \mathrm{v}_h)\Phi$. Assume that m' and m'' satisfy the hypothesis of the lemma with respect to Ω. If a_j occurs in Φ, it does so in Ω whence $\overline{\overline{m}}'_{2j} = \overline{\overline{m}}''_{2j}$. If $Ind(\mathrm{v}_j, \Phi) = 2$ and $j \neq h$, then $Ind(\mathrm{v}_j, \Omega) = 2$ and hence $\overline{\overline{m}}'_{2j-1} = \overline{\overline{m}}''_{2j-1}$. It follows that if a is an arbitrary integer, then the integers $m^* = C_{2h-1}(m', a)$ and

$m^{**} = C_{2h-1}(m'', a)$ satisfy the assumption of the lemma with respect to Φ. By the inductive hypothesis we obtain therefore

$$\mathfrak{Stf}(\Phi, C_{2h-1}(m', a)) \equiv \mathfrak{Stf}(\Phi, C_{2h-1}(m'', a))$$

for every a. Since $Val(\Omega, m')$ (or $Val(\Omega, m'')$) is defined as the smallest a satisfying the left (or the right) hand side of this equivalence or as 1 if no such integer a exists, we infer that $Val(\Omega, m') = Val(\Omega, m'')$.

Lemma 2 is thus proved.

Lemma 3. *If φ is a functional form, Φ a matrix form, and if α is a functional form such that the bound variables which occur in α and have the indices 2 occur neither in φ nor in Φ, then*

$$(1) \quad Val(S(i, \alpha, \varphi), m) = Val(\varphi, C_{2i}(m, Val(\alpha, m))),$$

$$(2) \quad \mathfrak{Stf}(S(i, \alpha, \Phi), m) \equiv \mathfrak{Stf}(\Phi, C_{2i}(m, Val(\alpha, m))).$$

Proof. We apply again the method of induction. If φ is 1, then (1) is evident. If $\varphi = \mathrm{a}_j$ with $j \neq i$, then the left hand side of (1) is $\overline{\overline{m}}_{2j}$ and the right hand side is $W(p_{2j}, C_{2i}(m, Val(\alpha, m))) = \overline{\overline{m}}_{2j}$. If $\varphi = \mathrm{a}_i$, then the left hand side of (1) is $Val(\alpha, m)$ and the right side is $W(p_{2i}, C_{2i}(m, Val(\alpha, m))) = Val(\alpha, m)$.

If $\varphi = \mathrm{v}_j$, then the left hand side of (1) is $\overline{\overline{m}}_{2j-1}$ and the right hand side is $W(p_{2j-1}, C_{2i}(m, Val(\alpha, m))) = \overline{\overline{m}}_{2j-1}$. Hence (1) is satisfied for the case when φ is one of the simplest functional forms 1, a_j, v_j.

Assume that (1) holds for two functional forms φ_1 and φ_2 and put $\varphi = \varphi_1 \boldsymbol{+} \varphi_2$. Since $S(i, \alpha, \varphi) = S(i, \alpha, \varphi_1) \boldsymbol{+} S(i, \alpha, \varphi_2)$ we obtain by the inductive hypothesis

$$Val(S(i, \alpha, \varphi), m) = Val(S(i, \alpha, \varphi_1), m) + Val(S(i, \alpha, \varphi_2), m) =$$
$$= Val(\varphi_1, C_{2i}(m, Val(\alpha, m))) + Val(\varphi_2, C_{2i}(m, Val(\alpha, m))) = Val(\varphi_1 \boldsymbol{+}$$
$$\boldsymbol{+} \varphi_2, C_{2i}(m, Val(\alpha, m))) = Val(\varphi, C_{2i}(m, Val(\alpha, m))).$$

It follows from these equations that the lemma is true for the functional form $\varphi = \varphi_1 \boldsymbol{+} \varphi_2$, and we can show quite similarly that it is true also for the functional form $\varphi_1 \times \varphi_2$ and for the matrix form $\varphi_1 \approx \varphi_2$.

If we assume that (2) holds for the matrix forms Φ_1 and Φ_2, we can show by the same method as above that (2) is also valid for the matrix form $\Phi_1 \rightarrow \Phi_2$.

Finally, let us assume that (2) holds for a matrix form Φ and put $\varphi = (\mu \mathrm{v}_h)\Phi$. Since $S(i, a, \varphi) = (\mu \mathrm{v}_h)S(i, a, \Phi)$, the integer $Val(S(i, a, \varphi), m)$ is equal to the least a such that

$$\mathfrak{Stsf}(S(i, a, \Phi), C_{2h-1}(m, a)) \tag{3}$$

or to 1 if no a with the property (3) exists.

According to the inductive hypothesis (3) is equivalent to

$$\mathfrak{Stsf}(\Phi, C_{2i}(C_{2h-1}(m, a),\ Val(a,\ C_{2h-1}(m, a)))). \tag{4}$$

Hence, $Val(S(i, a, \varphi), m)$ is the smallest a for which (4) holds or 1 if there is no such a.

Observe now that the integer $m' = C_{2h-1}(m, a)$ satisfies the equations $\overline{\overline{m}}'_j = \overline{\overline{m}}_j$ for all $j \neq 2h - 1$. According to the assumption of the lemma the index of v_h in a is 1 (because v_h occurs in φ). Hence, by lemma 2, $Val(a, m') = Val(a, m)$ and (4) is thus reduced to the equivalent formula

$$\mathfrak{Stsf}(\Phi, C_{2i}(C_{2h-1}(m, a),\ Val(a, m))).$$

Using the commutativity property of the function C established in the equation (1) of section 1 (p. 56) we transform the last formula into the following one

$$\mathfrak{Stsf}(\Phi, C_{2h-1}(C_{2i}(m,\ Val(a, m)), a)).$$

According to the definition given in section 2, the least a satisfying this formula is equal to

$$Val((\mu \mathrm{v}_h)\Phi, C_{2i}(m,\ Val(a, m))); \tag{5}$$

If no such a exists, then (5) is equal to 1. It follows that (5) and $Val(S(i, a, \varphi), m)$ are equal and the lemma 3 is proved.

Theorem 4. *If Φ is in $\mathfrak{Tr}$ and a in $\mathfrak{Re}$, then $S(i, a, \Phi)$ is in $\mathfrak{Tr}$.*

Proof. Let m be an arbitrary integer. Φ being an element of $\mathfrak{Tr}$, we have $\mathfrak{Stsf}(\Phi, C_{2i}(m,\ Val(a, m)))$ whence, by lemma 3,

$\mathfrak{Stsf}(S(i, a, \Phi), m)$. Since m is arbitrary, we obtain $S(i, a, \Phi) \in \mathfrak{Tr}$, q.e.d.

Theorem 4 shows that the class $\mathfrak{Tr}$ is closed under the operation of substitution.

Theorem 5. *If Φ is an axiom of* (S), *then Φ is in $\mathfrak{Tr}$.*

It will be sufficient to prove this only for the axioms (1) and (3) of the group II since the proofs for the remaining axioms are very easy and do not require any new technical device.

Let us assume that Φ is a matrix, that a_i occurs in Φ and that v_h does not occur in Φ. We have to show that if m is an integer such that $\mathfrak{Stsf}(\Phi, m)$, then

$$\mathfrak{Stsf}(S(i,(\mu v_h)S(i, v_h, \Phi), \Phi), m).$$

By lemma 3 this is equivalent to

$$\mathfrak{Stsf}(\Phi, C_{2i}(m, Val((\mu v_h)S(i, v_h, \Phi), m))). \tag{1}$$

First we calculate $Val((\mu v_h)S(i, v_h, \Phi), m)$. To find this integer we must look for the least a such that $\mathfrak{Stsf}(S(i, v_h, \Phi), C_{2h-1}(m, a))$, or what is the same

$$\mathfrak{Stsf}(\Phi, C_{2i}(C_{2h-1}(m, a), Val(v_h, C_{2h-1}(m, a)))). \tag{2}$$

Since $Val(v_h, C_{2h-1}(m, a)) = a$, (2) is equivalent to

$$\mathfrak{Stsf}(\Phi, C_{2i}(C_{2h-1}(m, a), a)). \tag{3}$$

Put $m' = C_{2h-1}(m, a)$, $m'' = C_{2i}(m', a)$. Since v_h does not occur in Φ we have by lemma 2

$$\mathfrak{Stsf}(\Phi, m) \equiv \mathfrak{Stsf}(\Phi, m')$$

and

$$\mathfrak{Stsf}(\Phi, C_{2i}(m, a)) \equiv \mathfrak{Stsf}(\Phi, C_{2i}(m', a)) \equiv \mathfrak{Stsf}(\Phi, m'')$$

which proves that (3) is equivalent to

$$\mathfrak{Stsf}(\Phi, C_{2i}(m, a)). \tag{4}$$

There exists at least one integer a satisfying the condition (4). Indeed, since $\mathfrak{Stsf}(\Phi, m)$ in virtue of our hypothesis made at the

beginning of the proof, and since $C_{2i}(m, \overline{\overline{m}}_{2i}) = m$, we infer that the integer $\overline{\overline{m}}_{2i}$ satisfies the condition (4).

Let a_0 be the least integer satisfying (4). It follows that

$$\mathfrak{Stz\!f}(\Phi, C_{2i}(m, a_0)), \tag{5}$$

$$Val((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), m) = a_0, \tag{6}$$

and these equations prove that the formula (1) is satisfied. Hence, the axiom II 1 is in $\mathfrak{Tr}$ (more exactly: all substitution-instances of this axiom are in $\mathfrak{Tr}$).

We pass now to the axiom II 3. Let us assume again that Φ is in $\mathfrak{M}$, that a_i occurs in Φ, and that v_h, a_k, a_l do not occur in Φ. Assume further that

$$\mathfrak{Stz\!f}((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi) \approx \mathrm{a}_k \,\boldsymbol{+}\, \mathrm{a}_l, m).$$

It follows easily from this assumption that

$$Val((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), m) = Val(\mathrm{a}_k, m) + Val(\mathrm{a}_l, m)$$

whence

$$Val((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), m) \neq 1. \tag{7}$$

The left hand side was calculated above: it is equal to the least a for which (4) is satisfied or to 1 if no such a exists. Formula (7) proves therefore that an a satisfying (4) exists. Let us denote the least a of this kind by a_0. It follows from the definition of a_0 that

$$a_0 > 1,\ \mathfrak{Stz\!f}(\Phi, C_{2i}(m, a_0)),\ \sim \mathfrak{Stz\!f}(\Phi, C_{2i}(m, b)) \text{ for } b < a_0. \tag{8}$$

Since

$$\mathfrak{Stz\!f}(S(i, \mathrm{a}_k, \Phi), m) \equiv \mathfrak{Stz\!f}(\Phi, C_{2i}(m, Val(\mathrm{a}_k, m)))$$

according to lemma 3 and since $Val(\mathrm{a}_k, m) < a_0$, we obtain from (8)

$$\sim \mathfrak{Stz\!f}(S(i, \mathrm{a}_k, \Phi), m)$$

i.e.,

$$\mathfrak{Stz\!f}(S(i, \mathrm{a}_k, \Phi) \rightarrow 1 \approx 1 \,\boldsymbol{+}\, 1, m).$$

This proves that axiom II 3 is in $\mathfrak{Tr}$ (more exactly that all substitution instances of that axiom are in $\mathfrak{Tr}$).

From theorems 1, 4, and 5 we obtain easily

Theorem 6. *All provable matrices are in* $\mathfrak{Tr}$.

Proof. Let Φ be a provable matrix and

$$\Phi_1, \Phi_2, \ldots, \Phi_n = \Phi$$

its formal proof. We shall show by induction that every Φ_i is in $\mathfrak{Tr}$. This is evident if Φ_i is in $\mathfrak{Ax}$ and in particular if $i = 1$ (since the first term of an arbitrary formal proof is always an axiom).

Let us assume that $j \leqslant n$ and that $\Phi_i \in \mathfrak{Tr}$ for $i < j$. Three cases are possible:

(1) *Φ_j is an axiom,*
(2) *there are k, l both less than j such that $\Phi_j = Mp(\Phi_k, \Phi_l)$,*
(3) *there are integers k, h, φ such that $k < j$, $\varphi \in \mathfrak{Re}$, and $\Phi_j = S(h, \varphi, \Phi_k)$.*

In each of these cases Φ_j is in $\mathfrak{Tr}$: in the case (1) in virtue of theorem 5, in the case (2) in virtue of theorem 1, and in the case (3) in virtue of theorem 4.

It follows now by induction that Φ_j is in $\mathfrak{Tr}$ for every j. Putting $j = n$ we obtain therefore the desired result.

Theorem 7. *The set $\mathfrak{Tr} \wedge \mathfrak{M}$ is consistent.*

Proof. If this set were inconsistent, every matrix and in particular the sentence $1 \approx 1 + 1$ would be $\mathfrak{Tr}$-provable. This is impossible since it would imply that $1 \approx 1 + 1$ is in $\mathfrak{Tr}$ whereas we know that for every $m \sim \mathfrak{Stsf}(1 \approx 1 + 1, m)$.

As a rather important corollary we obtain from theorem 7 the following result:

Theorem 8. *The set $\mathfrak{T}$ is consistent.*

Indeed, $\mathfrak{T}$ is a subset of $\mathfrak{Tr} \wedge \mathfrak{M}$ and the subset of a consistent set is itself consistent.

Theorem 9. *The set $\mathfrak{Tr} \wedge \mathfrak{M}$ is complete.*

Proof. Let Φ be an arbitrary sentence. Since no a_j occurs in Φ and every v_j which occurs in Φ has therein the index 1, the assumptions of lemma 2 are satisfied for arbitrary m' and m''. It

follows that if there exists at least one integer m such that $\mathfrak{Stsf}(\Phi, m)$, then every integer satisfies this condition and hence Φ is in $\mathfrak{Tr} \wedge \mathfrak{M}$. If no m satisfies the condition $\mathfrak{Stsf}(\Phi, m)$, then for every m

$$\mathfrak{Stsf}(\Phi \rightarrow 1 \approx 1 + 1, m)$$

and hence $\Phi \rightarrow 1 \approx 1 + 1 \in \mathfrak{Tr} \wedge \mathfrak{M}$. Hence either Φ or $\backsim \Phi$ is in $\mathfrak{Tr} \wedge \mathfrak{M}$ which proves that this set is complete.

To finish this Chapter we discuss still the problem of the ω-consistency of the set $\mathfrak{Tr} \wedge \mathfrak{M}$. We need the following auxiliary theorem:

Theorem 10. *If Φ is a matrix in which exactly one free variable a_i occurs, and if v_h is a bound variable which does not occur in Φ, then $Val((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), m)$ is the least p such that $S(i, D_p, \Phi)$ is in $\mathfrak{Tr}$ provided that such integers exist; otherwise $Val((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), m)$ is* 1.

Proof. Let us assume that

(1) $$S(i, D_p, \Phi) \textit{ is in } \mathfrak{Tr},$$

i.e., that $\mathfrak{Stsf}(S(i, D_p, \Phi), m)$ for an arbitrary m. According to lemma 3 this assumption is equivalent to the formula

$$\mathfrak{Stsf}(\Phi, C_{2i}(m, p)).$$

We shall show that this condition is in turn equivalent to

(2) $$\mathfrak{Stsf}(S(i, \mathrm{v}_h, \Phi), C_{2h-1}(m, p)).$$

Indeed, formula (2) is equivalent to $\mathfrak{Stsf}(\Phi, C_{2i}(C_{2h-1}(m, p), p))$, and since v_h does not occur in Φ, we may replace here $C_{2h-1}(m, p)$ by m without influencing the validity of the formula (cf. lemma 2).

We prove further that if (2) holds for at least one m, it does so for any m. Indeed, $S(i, \mathrm{v}_h, \Phi)$ is an expression in which only the variable v_h has the index 2 and in which no free variable occurs. According to lemma 2 formula (2) is equivalent to

$$\mathfrak{Stsf}(S(i, \mathrm{v}_h, \Phi), m')$$

where m' is an arbitrary integer such that

$$\overline{\overline{m}}'_{2h-1} = W(p_{2h-1}, C_{2h-1}(m, p)) = p.$$

In particular, we can take as m' the integer $C_{2h-1}(n, p)$ where n is wholly arbitrary since this choice of m' satisfies the above equation. Hence, if (2) holds for at least one m, it does so for every m.

We have thus shown that (1) is equivalent to each of the following conditions: (2) holds for at least one m; (2) holds for every m.

We can now prove theorem 10. Assume that there are integers satisfying (1) and let p be the least of them. Let m be arbitrary. We have then the formula (2) and for no $q < p$ can the formula

$$\mathfrak{Stz\mathfrak{f}}(S(i, \mathrm{v}_h, \Phi), C_{2h-1}(m, q))$$

be satisfied. Indeed, if this formula were true $S(i, D_q, \Phi)$ would be in $\mathfrak{Tr}$, and this contradicts the definition of p. Hence p is the least integer for which (2) holds which proves that

$$Val((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), m) = p.$$

Assume now that (1) does not hold for any p. Hence (2) is false for arbitrary p and m, and we obtain according to definition of *Val* (section 3, equation (8), p. 59)

$$Val((\mu \mathrm{v}_h)S(i, \mathrm{v}_h, \Phi), m) = 1$$

for arbitrary m.

Theorem 10 is thus proved.

As an easy corollary we obtain

Theorem 11. *The set* $\mathfrak{Tr} \wedge \mathfrak{M}$ *is* ω*-consistent.*

Proof. Let Φ be a matrix in which exactly one free variable a_1 occurs. Assume that $S(1, D_n, \sim\Phi)$ is in $\mathfrak{Tr}$ for $n = 1, 2, \ldots$. It follows that for no n $S(1, D_n, \Phi)$ is in $\mathfrak{Tr}$ and hence

$$Val((\mu \mathrm{v}_h)S(1, \mathrm{v}_h, \Phi), m) = 1$$

which proves that

(1) $$(\mu \mathrm{v}_h)S(1, \mathrm{v}_h, \Phi) \approx D_1 \textit{ is in } \mathfrak{Tr}.$$

Using theorem 1 of Chapter III, section 4, p. 45 we obtain that the following sentence

$$(\mu \mathrm{v}_h)S(1, \mathrm{v}_h, \Phi) \approx D_1 \rightarrow [S(1, (\mu \mathrm{v}_h)S(1, \mathrm{v}_h, \Phi), \Phi) \rightarrow S(1, D_1, \Phi)]$$

is in $\mathfrak{Tr}$. On account of (1) we obtain therefore that

$$S(1, (\mu \mathrm{v}_h) S(1, \mathrm{v}_h, \Phi), \Phi) \rightarrow S(1, D_1, \Phi)$$

is in $\mathfrak{Tr}$. This proves that $S(1, (\mu \mathrm{v}_h) S(1, \mathrm{v}_h, \Phi), \Phi)$ is not in $\mathfrak{Tr}$ since otherwise $S(1, D_1, \Phi)$ would be in $\mathfrak{Tr}$ which is not the case, as we have shown above.

Theorem 11 is thus proved. It implies, of course, that also *the set $\mathfrak{T}$ is ω-consistent.*

We note still the following corollaries to the theorem 10:

Theorem 12. *If Φ is a matrix in which exactly one free variable a_1 occurs and if v_h is a bound variable which does not occur in Φ, then*

$$(\mathrm{Ev}_h) S(i, \mathrm{v}_h, \Phi) \in \mathfrak{Tr} \equiv (\exists n) S(i, D_n, \Phi) \in \mathfrak{Tr},$$
$$(\mathrm{Av}_h) S(i, \mathrm{v}_h, \Phi) \in \mathfrak{Tr} \equiv (n) S(i, D_n, \Phi) \in \mathfrak{Tr}.$$

V

RECURSIVITY AND DEFINABILITY OF FUNCTIONS AND RELATIONS

1. The notion of $\mathfrak{K}$-definability. One of the chief problems of formal logic consists of the study of mutual relations between the mathematical entities (such as functions, sets, and so on) and expressions of a formal language. In our case we have to investigate relations existing between arithmetical notions on one side and matrices and numerical expressions on the other.

Relations which we have in mind are usually described by phrases such as "an expression is a formal definition of this or other arithmetical concept" or "an expression says in (S) that an arithmetical concept has this or other property". For instance the expression $\mathrm{a} \lessdot \mathrm{b}$ (cf. Chapter III, section 1, p. 43) is a formal definition of the "less-than" relation and the expression $(\mathrm{a} \times \mathrm{a}) \boldsymbol{+} 1$ is a formal definition of the function $x^2 + 1$. It is important to note that the formal definition of a given relation is not uniquely determined by the relation. For instance the matrices

$$\mathrm{a} \lessdot \mathrm{b},\ \backsim(\mathrm{b} \approx \mathrm{a})\ \&\ \backsim(\mathrm{b} \lessdot \mathrm{a}),\ \ (\mathrm{a} \lessdot \mathrm{b})\ \&\ [(\mu \mathrm{x})(\mathrm{x} \times \mathrm{x} \approx \mathrm{x}) \approx 1]$$

are formal definitions of one and the same relation.

The phrase "*an expression is a formal definition of a given arithmetical concept*" is certainly very vague and calls for explanation. Now it is not difficult to reduce this unclear notion to that of *truth* of a sentence. If we analyze the examples given above, we see easily that a numerical expression φ is a formal definition of a function F if as many free variables occur in φ as there are arguments in F, and if the sentence $\varphi(D_{n_1}, D_{n_2}, \ldots, D_{n_k}) \approx \approx D(F(n_1, n_2, \ldots, n_k))$ is true. Similarly, we shall say that a matrix Φ is a formal definition of a relation $\mathfrak{R}$ if as many free variables occur in Φ as there are arguments in $\mathfrak{R}$, and if the

sentence $\Phi(D_{n_1}, D_{n_2}, \ldots, D_{n_k})$ is true or false according as $\mathfrak{R}(n_1, n_2, \ldots, n_k)$ holds or not.

We see thus that the notion which interests us here depends on the meaning which we attach to the notion of a true sentence. Now we learned in the previous Chapters two notions which can be taken for the notion of "truth", namely that of a provable sentence and that of a sentence of the class $\mathfrak{Tr}$. It follows that we have at least two non-equivalent notions of a formal definability one of which corresponds to the notion of a provable sentence and the other to the notion of a sentence of the class $\mathfrak{Tr}$.

It will be convenient to take at present an arbitrary class $\mathfrak{K}$ of matrices as the class containing all "true" sentences. We shall assume only that this class is consistent and closed (cf. p. 41). In this way we obtain the notion of a $\mathfrak{K}$-definability of arithmetical concepts. Afterwards we shall specialize the class $\mathfrak{K}$ and shall put $\mathfrak{K} = \mathfrak{T}$ or $\mathfrak{K} = \mathfrak{Tr}$.

After these explanations it is easy to formulate the exact definition of the notion of formal definability.

Let F be a function with k arguments and $\mathfrak{R}$ a k termed relation. Further let $\mathfrak{K}$ be a closed class of matrices.

Definition 1. *The function F is called $\mathfrak{K}$-definable if there is a numerical expression φ in which exactly the k free variables* $\mathrm{a}_1, \ldots, \mathrm{a}_k$ *occur such that for arbitrary integers $n_1, \ldots, n_k$ the sentence $\varphi(D_{n_1}, \ldots, D_{n_k}) \approx D(F(n_1, \ldots, n_k))$ is in $\mathfrak{K}$.*

Each φ satisfying these conditions is said to be *$\mathfrak{K}$-associated with F.*

Definition 2. *The relation $\mathfrak{R}$ is called $\mathfrak{K}$-definable if there is a matrix Φ in which exactly the k free variables* $\mathrm{a}_1, \ldots, \mathrm{a}_k$ *occur such that for arbitrary integers $n_1, \ldots, n_k$ the following two conditions are satisfied:*

If $\mathfrak{R}(n_1, \ldots, n_k)$, then $\Phi(D_{n_1}, \ldots, D_{n_k})$ is in $\mathfrak{K}$,

If $\sim\mathfrak{R}(n_1, \ldots, n_k)$, then $\neg\Phi\ (D_{n_1}, \ldots, D_{n_k})$ is in $\mathfrak{K}$.

Each Φ satisfying these conditions is said to be *$\mathfrak{K}$-associated with $\mathfrak{R}$.*

If $\mathfrak{K} = \mathfrak{T}$ we shall say "*recursive*" instead of "*$\mathfrak{K}$-definable*".

Numerical expressions or matrices which are $\mathfrak{T}$-associated with functions or relations are called their *recursive definitions.*

If $\mathfrak{R} = \mathfrak{Tr}$ we shall say simply *definable* (or *definable in* (S)) instead of "$\mathfrak{R}$-definable". Numerical expressions or matrices which are $\mathfrak{Tr}$-associated with functions or relations are called their formal definitions [1].

We note an essential difference between the notions of recursivity and definability. Let us consider a formal system (S′) of arithmetic different from (S) but built essentially along the same lines. We can of course define the notions of recursivity and of definability with respect to this new system and it is natural to ask whether a function (or relation) which was recursive or definable with respect to the old system will remain to be recursive or definable with respect to the new system. It turns out that functions or relations definable in (S) are in general not definable in (S′), whereas functions and relations which possess recursive definitions in (S) will continue to do so in (S′). The notion of recursiveness is thus independent from the choice of the formal system whereas the notion of definability depends very essentially on the system, and changes its meaning when we pass from one system to another [2].

2. General properties of $\mathfrak{R}$-definable functions and relations. Throughout the whole section we assume that $\mathfrak{R}$ is a closed and consistent set of matrices.

Theorem 1. *Recursive functions and relations are $\mathfrak{R}$-definable.*

Proof. For every $\mathfrak{R}$ the set $\mathfrak{T}$ is contained in $\mathfrak{R}$.

Theorem 2. *The identity relation and the less than relation are recursive; further, so are the functions $m + n$, $m \cdot n$, and the function U of p arguments whose value is constantly* 1.

Proof. From the axiom $a \approx a$ and the corollary III 6 3 we infer that the matrix $a \approx b$ is a recursive definition of the identity

[1] The notion of recursiveness is due to Gödel [10]. It is equivalent to the notions of λ-definability, general recursivity, and computability introduced by Church [2], Kleene [12], and Turing [24]. The notion of definability is due to Tarski [21].

[2] Cf. Gödel [10].

relation. From theorems III 7 4 and III 7 8 it follows likewise that the matrix $\mathrm{a} \lessdot \mathrm{b}$ is a recursive definition of the relation $<$. From theorem III 6 1 we further obtain that numerical expressions $\mathrm{a} \mathbin{\pmb{+}} \mathrm{b}$ and $\mathrm{a} \times \mathrm{b}$ are recursive definitions of the functions $m + n$ and $m \cdot n$. Finally, we can show easily that the numerical expression $(\mu \mathrm{x})[(\mathrm{x} \approx 1) \,\&\, (\mathrm{a}_1 \approx \mathrm{a}_1) \,\&\, \ldots \,\&\, (\mathrm{a}_p \approx \mathrm{a}_p)]$ is a recursive definition of the function U.

Theorem 3. *If a function F and a relation $\mathfrak{R}$ are $\mathfrak{K}$-definable, then so are the function F' and relation $\mathfrak{R}'$ obtained from F and $\mathfrak{R}$ by an identification of a pair of arguments.*

Proof. Let φ be a $\mathfrak{K}$-definition of F and assume that F' arises from F by the identification of the i-th and the j-th argument $(i < j)$. Put

$$\varphi' = \varphi(\mathrm{a}_1, \ldots, \mathrm{a}_{j-1}, \mathrm{a}_i, \mathrm{a}_j, \ldots, \mathrm{a}_{k-1}).$$

Using theorem II 3 i we obtain

$$\varphi'(D_{n_1}, \ldots, D_{n_{k-1}}) = \varphi(D_{n_1}, \ldots, D_{n_{j-1}}, D_{n_i} D_{n_j}, \ldots, D_{n_{k-1}}).$$

Denoting the right hand side by $\overline{\varphi}$ we obtain therefore

$$\overline{\varphi} \approx D(F(n_1, \ldots, n_{j-1}, n_i, n_j, \ldots, n_{k-1})) \text{ is in } \mathfrak{K}$$

whence

$$\varphi'(D_{n_1}, \ldots, D_{n_{k-1}}) \approx D(F'(n_1, \ldots, n_{k-1})) \text{ is in } \mathfrak{K}.$$

This proves that φ' is $\mathfrak{K}$-associated with F'.

The proof for the relation $\mathfrak{R}'$ is similar.

Theorem 4. *If $\mathfrak{R}$ is a $\mathfrak{K}$-definable relation with l arguments and F and G are $\mathfrak{K}$-definable functions with k and l arguments, then the relation*

$$(1) \qquad \mathfrak{R}(n_1, \ldots, n_{j-1}, F(m_1, \ldots, m_k), n_{j+1}, \ldots, n_l)$$

and the function

$$(2) \qquad G(n_1, \ldots, n_{j-1}, F(m_1, \ldots, m_k), n_{j+1}, \ldots, n_l)$$

are $\mathfrak{K}$-definable.

Proof. Φ, φ, ψ be $\mathfrak{K}$-definitions of $\mathfrak{R}$, F, and G. We maintain that the matrix

$$\Psi = \Phi(\mathrm{a}_1, \ldots, \mathrm{a}_{j-1}, \varphi(\mathrm{a}_l, \ldots, \mathrm{a}_{l+k-1}), \mathrm{a}_j, \ldots, \mathrm{a}_{l-1})$$

is a $\mathfrak{K}$-definition of the relation (1) and the numerical expression

$$\vartheta = \psi(\mathrm{a}_1, \ldots, \mathrm{a}_{j-1}, \varphi(\mathrm{a}_l, \ldots, \mathrm{a}_{l+k-1}), \mathrm{a}_j, \ldots, \mathrm{a}_{l-1})$$

is a $\mathfrak{K}$-definition of the function (2).

It will be sufficient to consider only the case of the function (2) since the proof of the other part of the theorem is entirely similar.

Let $n_1, \ldots, n_l, m_1, \ldots, m_k$ be arbitrary integers and p the value of the function (2) for these arguments. Further let

$$q = F(m_1, \ldots, m_k).$$

Since $G(n_1, \ldots, n_{j-1}, q, n_{j+1}, \ldots, n_l) = p$ we obtain from the assumption that ψ is a $\mathfrak{K}$-definition of G the formula

(3) $\qquad \psi(D_{n_1}, \ldots, D_{n_{j-1}}, D_q, D_{n_{j+1}}, \ldots, D_{n_l}) \approx D_p$ *is in* $\mathfrak{K}$.

Since φ is a $\mathfrak{K}$-definition of F, we obtain similarly

(4) $\qquad \varphi(D_{m_1}, \ldots, D_{m_k}) \approx D_q$ *is in* $\mathfrak{K}$.

Observe now that by theorem II 4 1

$$\mathrm{a}_h \approx \mathrm{a}_j \rightarrow S(j, \mathrm{a}_h, \psi) \approx \psi \textit{ is in } \mathfrak{T}$$

where h is any integer such that a_h occurs neither in ψ nor in φ. Let us perform the following substitutions in this formula:

$$D_{n_1}, \ldots, D_{n_{j-1}}, D_{n_{j+1}} \ldots, D_{n_l}$$

are substituted for $\mathrm{a}_1, \ldots, \mathrm{a}_{j-1}, \mathrm{a}_{j+1}, \ldots, \mathrm{a}_l$, D_q is substituted for a_j, and $\varphi(D_{m_1}, \ldots, D_{m_k})$ for a_h. Using modus ponens and the formula (4) we then obtain

$$\psi(D_{n_1}, \ldots, D_{n_{j-1}}, \varphi(D_{m_1}, \ldots, D_{m_k}), D_{n_{j+1}}, \ldots, D_{n_l}) \approx$$
$$\approx \psi(D_{n_1}, \ldots, D_{n_{j-1}}, D_q, D_{n_{j+1}}, \ldots, D_{n_l}) \in \mathfrak{K}$$

whence by (3) and the definition of ϑ

$$\vartheta(D_{n_1}, \ldots, D_{n_{j-1}}, D_{n_{j+1}}, \ldots, D_{n_l}, D_{m_1}, \ldots, D_{m_k}) \approx D_p \in \mathfrak{K}.$$

This proves that ϑ is a $\mathfrak{K}$-definition of the function (2), q.e.d.

Theorem 5. *If relations $\mathfrak{R}$ and $\mathfrak{S}$ are $\mathfrak{K}$-definable, then so are the relations $\mathfrak{R} \vee \mathfrak{S}$ and $-\mathfrak{R}$.*

Proof. Let $\mathfrak{R}$ be a k-termed and $\mathfrak{S}$ an l-termed relation and let Φ and Ψ be their $\mathfrak{K}$-definitions.

We shall show that $\backsim\Phi$ is a $\mathfrak{K}$-definition of $-\mathfrak{R}$. Indeed, if $-\mathfrak{R}(n_1, \ldots, n_k)$, then $\sim\mathfrak{R}(n_1, \ldots, n_k)$ and hence by the definition 2 $\backsim \Phi(D_{n_1}, \ldots, D_{n_k}) \in \mathfrak{K}$. Conversely, if $\sim -\mathfrak{R}(n_1, \ldots, n_k)$, then $\mathfrak{R}(n_1, \ldots, n_k)$ and hence $\Phi(D_{n_1}, \ldots, D_{n_k}) \in \mathfrak{K}$. Since $\Phi \rightarrow \backsim\backsim\Phi \in \mathfrak{T}$, we obtain that $\backsim\backsim\Phi(D_{n_1}, \ldots, D_{n_k}) \in \mathfrak{K}$.

It can be shown similarly that the matrix

$$[\backsim \Phi(a_1, \ldots, a_k)] \rightarrow \Psi(a_{k+1}, \ldots, a_{k+l})$$

is a $\mathfrak{K}$-definition of the relation $\mathfrak{R} \vee \mathfrak{S}$. Theorem 5 is thus proved.

Theorem 6. *If $\mathfrak{R}$ is a k-termed $\mathfrak{K}$-definable relation such that*

$$(*) \qquad (n_1) \ldots (n_{j-1})(n_{j+1}) \ldots (n_k)(\exists n_j)\mathfrak{R}(n_1, \ldots, n_k),$$

then the function $F = \min_j \mathfrak{R}$ is $\mathfrak{K}$-definable.

Proof. Let Φ be a $\mathfrak{K}$-definition of $\mathfrak{R}$ and put

$$\varphi = (\mu \mathrm{v}_h)\Phi(\mathrm{a}_1, \ldots, \mathrm{a}_{j-1}, \mathrm{v}_h, \mathrm{a}_j, \ldots, \mathrm{a}_{k-1})$$

where h is an integer such that v_h does not occur in Φ. We shall show that φ is a $\mathfrak{K}$-definition of F.

Let $n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_k$ be arbitrary integers and denote $F(n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_k)$ by p. According to (*) p is the least integer such that

$$\mathfrak{R}(n_1, \ldots, n_{j-1}, p, n_{j+1}, \ldots, n_k).$$

Denote by the bar the operation of substitution of digits $D_{n_1}, \ldots, D_{n_{j-1}}, D_{n_{j+1}}, \ldots, D_{n_k}$ for the variables

$$\mathrm{a}_1, \ldots, \mathrm{a}_{j-1}, \mathrm{a}_{j+1}, \ldots, \mathrm{a}_k.$$

By the definition of Φ we obtain the formulas

$$S(j, D_p, \bar{\Phi}) \in \mathfrak{K},$$
$$S(j, D_i, \backsim\bar{\Phi}) \in \mathfrak{K} \text{ for } i = 1, 2, \ldots, p-1.$$

Using theorem III 7 12 we obtain from these formulas

$$(\mu \mathrm{v}_h) S(j, \mathrm{v}_h, \overline{\Phi}) \approx D_p \in \mathfrak{K}$$

which proves that $\overline{\varphi} \approx D_p \in \mathfrak{K}$. Hence φ is a $\mathfrak{K}$-definition of F which proves the theorem.

We shall see later that theorem 6 is in general false when $\mathfrak{R}$ does not satisfy the condition (*). In the particular case $\mathfrak{K} = \mathfrak{Tr}$ the condition (*) can, however, be dispensed with. This is shown in the next theorem:

Theorem 7. *If $\mathfrak{R}$ is a definable k-termed relation and $1 \leqslant j \leqslant k$, then the function $F = \min_j \mathfrak{R}$ is definable. Moreover, if Φ is a formal definition of $\mathfrak{R}$, then $\varphi = (\mu \mathrm{v}_h) S(j, \mathrm{v}_h, \Phi)$ is a formal definition of F* (v_h is an arbitrary bound variable which does not occur in Φ).

Proof. Let $n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_k$ be arbitrary integers and let the bar over an expression denote as before the operation of substitution of the digits $D_{n_1}, \ldots, D_{n_{j-1}}, D_{n_{j+1}}, \ldots, D_{n_k}$ for the variables $\mathrm{a}_1, \ldots, \mathrm{a}_{j-1}, \mathrm{a}_{j+1}, \ldots, \mathrm{a}_k$. Put

$$p = F(n_1, \ldots, n_{j-1}, n_{j+1}, \ldots, n_k).$$

Hence p is either the least integer such that

$$\mathfrak{R}(n_1, \ldots, n_{j-1}, p, n_{j+1}, \ldots, n_k),$$

or 1 if no such integer exists.

Φ being a formal definition of $\mathfrak{R}$, we obtain in the first case

$$p \textit{ is the least integer such that } S(j, D_p, \overline{\Phi}) \in \mathfrak{Tr},$$

and in the second

$$p = 1.$$

Using theorem IV 4 10 we obtain therefore

$$Val((\mu \mathrm{v}_h) S(i, \mathrm{v}_h, \overline{\Phi}), m) = p$$

for an arbitrary m. This proves that $\mathfrak{Stf}(\overline{\varphi} \approx D_p, m)$ for every m and hence $\varphi \approx D_p$ is in $\mathfrak{Tr}$.

Theorem 7 is thus proved. As an easy corollary we obtain

Corollary 8. *If Φ is a formal definition of a k-termed relation $\mathfrak{R}$, then* $(\mathrm{Ev}_h)S(j, \mathrm{v}_h, \Phi)$ *is a formal definition of the relation*

$$(\exists n_j)\mathfrak{R}(n_1, \ldots, n_k).$$

In order to explain the importance of theorems proved in this section, we make the following observations: We have learned in Chapter I (section **4**, p. 19) a series of operations which enable us to construct new relations or functions from other given relations or functions. The most important of these operations are: (1) the Boolean operations on relations, (2) substitution and identification of variables, (3) the operation min which leads from a relation to a function, and (4) operations corresponding to bounded and unbounded quantifiers.

Theorems 1—7 show that the operations (1) and (2) performed on $\mathfrak{K}$-definable relations or functions yield again $\mathfrak{K}$-definable relations or functions. This is still true for the operation (3) provided that we apply it to relations which satisfy the condition (*) of theorem 6. This implies, of course, that also the operations corresponding to the bounded quantifiers preserve $\mathfrak{K}$-definability of relations (cf. Chapter I, section **4**, p. 21). The operation min applied to an arbitrary $\mathfrak{K}$-definable relation yield in general a function which is not $\mathfrak{K}$-definable; also the operations corresponding to the unbounded quantifiers do not preserve the property of $\mathfrak{K}$-definability. In the special case $\mathfrak{K} = \mathfrak{Tr}$ however the class of $\mathfrak{K}$-definable functions and relations is closed also with respect to these operations.

Applying these general theorems we shall solve in the next section the problem of $\mathfrak{K}$-definability of the various functions and relations which we have introduced in the previous Chapters.

3. Recursivity of functions and relations corresponding to the syntactical notions. We begin with theorems in which the recursivity of the auxiliary functions defined in Chapter I is established.

Theorem 1. *The functions J, J_3, K_1, K_2, L_1, L_2, L_3 are recursive.*

Proof. $J(i, j)$ can be defined as the unique x satisfying the relation $2x = (i + j - 1)\cdot(i + j - 2) + 2j$. Since this relation is recursive and satisfies the condition (*) of theorem 2 6 (cf. p. 78), we obtain the result that the function $J(i, j) = \min_x[2x = (i+j-1)\cdot(i + j - 2) + 2j]$ is recursive. From this follows the recursivity of the function $J_3(i, j, k) = J(i, J(j, k))$ (cf. 2, theorem 4, p. 76). The recursivity of the functions K_1 and K_2 results from theorem 2 6 and the formulas

$$x = K_1(n) \equiv (\exists y)_n\, J(x, y) = n,$$
$$x = K_2(n) \equiv (\exists y)_n\, J(y, x) = n.$$

The recursivity of the functions L_1, L_2, L_3 can be proved similarly.

Theorem 2. *The functions* R, $\bar{g}_i$, $\bar{g}_*$ (cf. Chapter I, section 2, pp. 15 and 17) *are recursive.*

Proof. $R(m, n)$ can be defined as the unique (and hence the least) x satisfying the following recursive relation

$$(m < n)\cdot(x = 1 + m) \vee (n \leqslant m)\cdot(\exists y)_m(m = ny + x - 1).$$

The recursivity of the functions $\bar{g}_i$ and $\bar{g}_*$ follows from the fact that these functions result by substitutions from the functions R, L_1, L_2, L_3, $x + y$, and xy.

Theorem 3. *Let G and H be two $\mathfrak{R}$-definable functions of one and three arguments. The function F obtained from G and H by the schema of the primitive recursion* (cf. Chapter I, section 5, p. 21) *is $\mathfrak{R}$-definable.*

Proof. The theorem results immediately from the explicit definition of the function F, given in Chapter I, section 5, and the remark that in this definition the min-operator has been applied only to relations which satisfy the condition (*) of theorem 2 6.

Corollary 4. *The function a^b is recursive.*

Proof. This function can be obtained by the schema of the primitive recursion from recursive functions (namely $G(y) = y$ and $H(x, y, z) = yz$).

Corollary 5. *The function* $T(a)$ (cf. Chapter I, section 2, theorem 3) *is recursive.*

Proof. $T(a)$ can be obtained from recursive functions by the operation of substitution.

Theorem 6. *The classes* $\mathfrak{Ex}$, $\mathfrak{Mf}$, $\mathfrak{Ff}$, $\mathfrak{M}$, $\mathfrak{Re}$, *and the functions* Oc, Ind, *and* S *are recursive.*

Proof. This follows immediately from the explicit definition of these classes and functions which has been given in Chapter II, section **4**, p. 37. Note that only bounded quantifiers have been used in this definition.

Theorem 7. *The set of axioms of* (S) *is recursive.*

Proof. This follows from the explicit definition of this set given in Chapter II, section **5**.

Theorem 8. *The function* Mp *is recursive.*

Proof. This follows from the definition of this function given in Chapter II, section **6**, p. 40.

Theorem 9. *If* $\mathfrak{L}$ *is a* $\mathfrak{R}$*-definable set of matrices, then the set* $\mathfrak{P}_{\mathfrak{L}}$ *of integers representing formal* $\mathfrak{L}$*-proofs is* $\mathfrak{R}$*-definable.*

Proof. This follows from the explicit definition of the set $\mathfrak{P}_{\mathfrak{L}}$, given in Chapter II, section **6**, p. 40.

Theorem 10. *If* $\mathfrak{L}$ *is a definable set of matrices, then the set* $\mathfrak{T}_{\mathfrak{L}}$ *is also definable.*

Proof. The set $\mathfrak{T}_{\mathfrak{L}}$ has been defined in Chapter II, section **6**, p. 41 by means of the set $\mathfrak{P}_{\mathfrak{L}}$ and an unbounded existential quantifier. Hence the theorem follows by theorem 7 of section **2**, p. 79.

We shall see in Chapter VI (section **1**, corollary 3) that the set $\mathfrak{T}_{\mathfrak{L}}$ is not necessarily recursive even if the set $\mathfrak{L}$ is recursive. Hence we cannot, in general, replace the word "definable" in theorem 10 by the word "recursive".

4. A theorem concerning the min-operator. In this section we shall prove the following theorem:

Theorem 1. *If* $\mathfrak{R}$ *is a closed consistent set satisfying the following condition*

(**) *if* $\mathfrak{R}$ *is a* $\mathfrak{K}$*-definable relation, then so is the function* $\min_j \mathfrak{R}$ *for every value of* j,

then every definable relation is $\mathfrak{K}$*-definable.*

Proof. Let φ be a numerical expression in which exclusively the free variables $a_{i_1}, \ldots, a_{i_n}$ occur. We denote by (φ, m) the numerical expression $\varphi(D(\overline{\overline{m}}_{i_1}), \ldots, D(\overline{\overline{m}}_{i_n}))$ and adopt a similar notation for matrices.

We shall show that for every numerical expression φ there is a numerical expression ψ in which the same free variables occur such that

$$(1)\qquad (\varphi, m) \approx D_p \in \mathfrak{Tr} \equiv (\psi, m) \approx D_p \in \mathfrak{K}.$$

We also show that for every matrix Φ there is a matrix Ψ in which the same free variables occur such that

$$(2)\qquad (\Phi, m) \in \mathfrak{Tr} \equiv (\Psi, m) \in \mathfrak{K},$$

$$(3)\qquad (\sim\Phi, m) \in \mathfrak{Tr} \equiv (\sim\Psi, m) \in \mathfrak{K}.$$

To show this, it is sufficient to construct the required numerical expression ψ for $\varphi = 1$, $\varphi = a_h$, and further to prove that if ψ and Ψ exist for two numerical expressions φ_1 and φ_2 and two matrices Φ_1 and Φ_2, then they also exist for the numerical expressions $\varphi_1 + \varphi_2$, $\varphi_1 \times \varphi_2$, $(\mu v_h)S(j, v_h, \Phi_1)$ and matrices $\varphi_1 \approx \varphi_2$, $\Phi_1 \rightarrow \Phi_2$.

In cases $\varphi = 1$, $\varphi = a_h$, it is sufficient to take $\psi = \varphi$.

Assume that φ_1, φ_2 are numerical expressions and that ψ_1, ψ_2 are numerical expressions in which the same free variables as in φ_1, φ_2 occur and which satisfy the conditions

$$(4)\qquad (\varphi_i, m) \approx D_p \in \mathfrak{Tr} \equiv (\psi_i, m) \approx D_p \in \mathfrak{K}\ (i = 1, 2).$$

We shall prove that

$$(5_1)\qquad (\varphi_1 + \varphi_2, m) \approx D_p \in \mathfrak{Tr} \equiv (\psi_1 + \psi_2, m) \approx D_p \in \mathfrak{K},$$

$$(5_2)\qquad (\varphi_1 \times \varphi_2, m) \approx D_p \in \mathfrak{Tr} \equiv (\psi_1 \times \psi_2, m) \approx D_p \in \mathfrak{K}.$$

Since these proofs are entirely similar, it will be sufficient to carry out only one of them, e.g. the first.

From the properties of the operation of *substitution* it follows that $(\varphi_1 \boldsymbol{+} \varphi_2, m) = (\varphi_1, m) \boldsymbol{+} (\varphi_2, m)$. The *expressions* (φ_i, m) being *numerals* possess well defined values $Val((\varphi_i, m), n) = q_i$ which are independent of n, and the sentence $(\varphi_1, m) \boldsymbol{+} (\varphi_2, m) \approx D_p$ is in $\mathfrak{Tr}$ if and only if $p = q_1 + q_2$. Furthermore, $(\varphi_i, m) \approx D_{q_i} \in \mathfrak{Tr}$ for $i = 1, 2$.

Using the assumption (4) we obtain

$$(\psi_i, m) \approx D_{q_i} \in \mathfrak{R} \quad (i = 1, 2)$$

whence

$$(\psi_1, m) \boldsymbol{+} (\psi_2, m) \approx D_{q_1} \boldsymbol{+} D_{q_2} \in \mathfrak{R},$$

i.e. (cf. Chapter III, section **6**, theorem 1, p. 49)

$$(\psi_1 \boldsymbol{+} \psi_2, m) \approx D_{q_1+q_2} \in \mathfrak{R}.$$

Assume now that $(\varphi_1 \boldsymbol{+} \varphi_2, m) \approx D_p \in \mathfrak{Tr}$, i.e. that $p = q_1 + q_2$. It follows that $(\psi_1 \boldsymbol{+} \psi_2, m) \approx D_p \in \mathfrak{R}$. Conversely, if

$$(\varphi_1 \boldsymbol{+} \varphi_2, m) \approx D_p$$

is not in $\mathfrak{Tr}$, then $q_1 + q_2 \neq p$. If $(\psi_1 \boldsymbol{+} \psi_2, m) \approx D_p$ were in $\mathfrak{R}$, we would obtain $D_p \approx D_{q_1+q_2} \in \mathfrak{R}$ and this implies the inconsistency of $\mathfrak{R}$ (cf. Chapter III, section **6**, corollary 3, p. 50). Hence

$$(\psi_1 \boldsymbol{+} \psi_2, m) \approx D_p$$

is not in $\mathfrak{R}$. This completes the proof of the formula (5_1).

We show further that under the assumption (4)

$$(\varphi_1 \approx \varphi_2, m) \in \mathfrak{Tr} \equiv (\psi_1 \approx \psi_2, m) \in \mathfrak{R}, \tag{6}$$

$$(\sim\varphi_1 \approx \varphi_2, m) \in \mathfrak{Tr} \equiv (\sim\psi_1 \approx \psi_2, m) \in \mathfrak{R}. \tag{7}$$

Indeed, $(\varphi_1 \approx \varphi_2, m)$ is in $\mathfrak{Tr}$ if and only if $q_1 = q_2$. Since $(\psi_i, m) \approx D_{q_i}$ is in $\mathfrak{R}$ for $i = 1, 2$, we infer that

$$(\psi_1, m) \approx (\psi_2, m) \leftrightarrow D_{q_1} \approx D_{q_2}$$

is in $\mathfrak{R}$.

Hence, if $q_1 = q_2$, the *sentence* $(\psi_1 \approx \psi_2, m)$ is in $\mathfrak{R}$; if $q_1 \neq q_2$, the *sentence* $(\sim\psi_1 \approx \psi_2, m)$ is in $\mathfrak{R}$. The equivalences (6) and (7) are thus proved.

Let us assume that Φ_1 and Φ_2 are matrices, and that

(8) $$(\Phi_i, m) \in \mathfrak{Tr} \equiv (\Psi_i, m) \in \mathfrak{R},$$

(9) $$(\backsim\Phi_i, m) \in \mathfrak{Tr} \equiv (\backsim\Psi_i, m) \in \mathfrak{R}$$

for $i = 1, 2$. Since $(\Phi_1 \rightarrow \Phi_2, m)$ is in $\mathfrak{Tr}$ if and only if either $(\backsim\Phi_1, m)$ or (Φ_2, m) is in $\mathfrak{Tr}$, we obtain easily that

$$(\Phi_1 \rightarrow \Phi_2, m) \in \mathfrak{Tr} \equiv (\Psi_1 \rightarrow \Psi_2, m) \in \mathfrak{R}.$$

It remains to consider the operation μ. Assume that (2) and (3) hold for matrices Φ and Ψ, and that the free variables which occur in these matrices are $\mathrm{a}_i, \mathrm{a}_j, \ldots, \mathrm{a}_h$. The relation

$$\mathfrak{N} = \lambda pq \ldots r[\Phi(D_p, D_q, \ldots, D_r) \in \mathfrak{Tr}]$$

is $\mathfrak{R}$-definable. Indeed, it follows from (2) and (3) that

$$\mathfrak{N}(p, q, \ldots, r) \supset \Psi(D_p, D_q, \ldots, D_r) \in \mathfrak{R},$$
$$\sim \mathfrak{N}(p, q, \ldots, r) \supset \backsim\Psi(D_p, D_q, \ldots, D_r) \in \mathfrak{R}.$$

According to the assumption (**) the function $F = \min_s \mathfrak{N}$ is $\mathfrak{R}$-definable, i.e. there exists a numerical expression ψ' in which exactly the free variables $\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_{n-1}$ occur and which satisfies the condition

(10) $$[\min_s \mathfrak{N}(p, q, \ldots, r) = t] \supset [\psi'(D_p, D_q, \ldots, D_r) \approx D_t \in \mathfrak{R}].$$

Assuming that a_k is the s-th of the variables $\mathrm{a}_i, \mathrm{a}_j, \ldots, \mathrm{a}_h$ and that v_l does not occur in Φ, we infer from theorem 1 7 that $\varphi = (\mu\mathrm{v}_l)S(k, \mathrm{v}_l, \Phi)$ is a formal definition of $\min_s \mathfrak{N}$, and hence that the left hand side of the formula (10) is equivalent to

$$\varphi(D_p, D_q, \ldots, D_r) \approx D_t \in \mathfrak{Tr}.$$

Formula (10) yields thus

(11) $$\varphi(D_p, D_q, \ldots, D_r) \approx D_t \in \mathfrak{Tr} \supset \psi'(D_p, D_q, \ldots, D_r) \approx D_t \in \mathfrak{R}.$$

If $\varphi(D_p, D_q, \ldots, D_r) \approx D_t$ is not in $\mathfrak{Tr}$ and u is the value of $\varphi(D_p, D_q, \ldots, D_r)$, then $\varphi(D_p, D_q, \ldots, D_r) \approx D_u$ is in $\mathfrak{Tr}$ and we obtain from (11)

$$\psi'(D_p, D_q, \ldots, D_r) \approx D_u \in \mathfrak{R}.$$

It follows that $\psi'(D_p, D_q, \ldots, D_r) \approx D_t$ cannot be in $\mathfrak{K}$ since otherwise $D_t \approx D_u$ would be in $\mathfrak{K}$ and $\mathfrak{K}$ would be inconsistent.

If we replace in this proof ψ' by the numerical expression $\psi = \psi'(\mathrm{a}_i, \mathrm{a}_j, \ldots, \mathrm{a}_h)$, we obtain

$$\varphi(D_p, D_q, \ldots, D_r) \approx D_t \in \mathfrak{Tr} \equiv \psi(D_p, D_q, \ldots, D_r) \approx D_t \in \mathfrak{K}$$

where ψ is a numerical expression in which the same free variables occur as in φ.

The proof of (1), (2), and (3) is thus complete.

The statement of theorem 1 follows now by the simple remark that every definable relation is equivalent to a relation of the form

$$\lambda pq \ldots r[\Phi(D_p, D_q, \ldots, D_r) \in \mathfrak{Tr}].$$

5. Recursively enumerable sets. We noted already in section 2 that the min-operator applied to a binary recursive relation yields a set which is, in general, not recursive. The same holds true for other operators defined in the terms of the min-operator and in particular for the existential quantifier. In other words if $\mathfrak{R}$ is a binary recursive relation, then the set

(1) $$\lambda m[(\exists n)\mathfrak{R}(m, n)]$$

is, in general, not recursive.

Sets of the form (1) are called *recursively enumerable.*

An important example of recursively enumerable sets is furnished by the following theorem:

Theorem 1. *If $\mathfrak{K}$ is a recursive set, then the set $\mathfrak{T}_{\mathfrak{K}}$* (cf. Chapter II, section 6, p. 40) *is recursively enumerable.*

Proof. It follows from the definition of $\mathfrak{T}_{\mathfrak{K}}$ that

$$\mathfrak{T}_{\mathfrak{K}} = \lambda m[(\exists n)\mathfrak{R}(m, n)], \text{ where } \mathfrak{R} = \lambda mn[(n \in \mathfrak{P}_{\mathfrak{K}}) \cdot (\bar{n}_* = m)].$$

Since $\mathfrak{R}$ is a recursive relation (cf. section 3, theorems 2 and 9) the set $\mathfrak{T}_{\mathfrak{K}}$ is recursively enumerable.

The theory of recursively enumerable sets has been developed very extensively [8]. We shall need only a few results from this theory.

[8] Cf. Post [16].

Theorem 2. *If the set $\mathfrak{A}$ and its complement $-\mathfrak{A}$ are recursively enumerable, then they are both recursive* [4].

Proof. It follows from the definition of recursively enumerable sets that there exist recursive binary relations $\mathfrak{R}$ and $\overline{\mathfrak{R}}$ such that

$$m \in \mathfrak{A} \equiv (\exists n)\mathfrak{R}(m, n), \quad m \in -\mathfrak{A} \equiv (\exists n)\overline{\mathfrak{R}}(m, n).$$

Hence, for every m there exists an n such that $\mathfrak{R}(m, n) \vee \overline{\mathfrak{R}}(m, n)$, and since the relation $\mathfrak{S} = \lambda mn[\overline{\mathfrak{R}}(m, n) \vee \mathfrak{R}(m, n)]$ is recursive, so is the function $F = \min_2 \mathfrak{S}$. It follows that the set $\lambda m[\mathfrak{R}(m, F(m))]$ is recursive, and since this set clearly coincides with $\mathfrak{A}$, we infer that $\mathfrak{A}$ and hence also its complement $-\mathfrak{A}$ is recursive.

Theorem 3. *Recursive sets are recursively enumerable.*

Proof. It is sufficient to remark that

$$m \in \mathfrak{A} \equiv (\exists n)[m \in \mathfrak{A} \cdot (n = n)],$$

and that the relation $\lambda mn[m \in \mathfrak{A} \cdot (n = n)]$ is recursive if the set $\mathfrak{A}$ is recursive.

Theorem 4. *The union of two or more recursively enumerable sets is recursively enumerable.*

Proof. The theorem follows at once from the well-known formula

$$(\exists n)\mathfrak{R}_1(m, n) \vee (\exists n)\mathfrak{R}_2(m, n) \equiv (\exists n)[\mathfrak{R}_1(m, n) \vee \mathfrak{R}_2(m, n)]$$

and the remark that if the relations $\mathfrak{R}_1$, $\mathfrak{R}_2$ are recursive, then so is their union $\mathfrak{R}_1 \vee \mathfrak{R}_2$.

[4] This theorem is due to Kleene [12], theorem V, p. 56.

VI

PROOFS OF INCOMPLETENESS THEOREMS

1. $\mathfrak{K}$-undefinability of the set $\mathfrak{T}_{\mathfrak{K}}$. We prove in this section the important theorem stating that if $\mathfrak{K}$ is a closed consistent set of matrices, then the set $\mathfrak{S} \wedge \mathfrak{K}$ is not $\mathfrak{K}$-definable. This theorem has many interesting applications, e.g. it provides us with examples of non-recursive sets and sets which are not definable in (S). We shall also see that the incompleteness theorems are easy corollaries to the theorem to be proved here. The method of proof is clearly connected with the well-known argument with the help of which Cantor showed that the set of all reals is not denumerable.

Theorem 1. *If $\mathfrak{K}$ is a closed consistent set of matrices, then the set $\mathfrak{S} \wedge \mathfrak{K}$ is not $\mathfrak{K}$-definable.*

Proof. Assume that the set $\mathfrak{S} \wedge \mathfrak{K}$ (i.e. the set of sentences which are elements of $\mathfrak{K}$) is $\mathfrak{K}$-definable, and consider the set $\lambda n[S(1, D_n, n) \text{ non} \in \mathfrak{S} \wedge \mathfrak{K}]$. According to theorems IV 2 4 and IV 2 5 this set is $\mathfrak{K}$-definable and hence there exists a matrix Φ in which exactly one free variable a_1 occurs and which satisfies the formulas

$$S(1, D_n, n) \text{ non} \in \mathfrak{S} \wedge \mathfrak{K} \supset \Phi(D_n) \in \mathfrak{K},$$
$$S(1, D_n, n) \in \mathfrak{S} \wedge \mathfrak{K} \supset \sim\Phi(D_n) \in \mathfrak{K}.$$

Remembering that $\Phi(D_n)$ is an abbreviation for $S(1, D_n, \Phi)$ and that this matrix is in $\mathfrak{S}$, we can rewrite these formulas as follows:

(1) $\qquad S(1, D_n, n) \text{ non} \in \mathfrak{S} \wedge \mathfrak{K} \supset S(1, D_n, \Phi) \in \mathfrak{S} \wedge \mathfrak{K},$

(2) $\qquad S(1, D_n, n) \in \mathfrak{S} \wedge \mathfrak{K} \supset \sim S(1, D_n, \Phi) \in \mathfrak{S} \wedge \mathfrak{K}.$

Put $n = \Phi$. From (1) we obtain then $S(1, D_\Phi, \Phi) \in \mathfrak{S} \wedge \mathfrak{K}$ and hence (2) yields $\sim S(1, D_\Phi, \Phi) \in \mathfrak{S} \wedge \mathfrak{K}$. This proves that the set $\mathfrak{K}$ is inconsistent.

It follows that our assumption was wrong and that the set $\mathfrak{S} \wedge \mathfrak{R}$ is not $\mathfrak{R}$-definable.

Theorem 1 is thus proved. Taking $\mathfrak{R} = \mathfrak{Tr}$ or $\mathfrak{R} = \mathfrak{T}$ we obtain the following corollaries:

Corollary 2. *The set $\mathfrak{Tr} \wedge \mathfrak{S}$ and hence also the set $\mathfrak{Tr}$ is not definable in* (S) [1].

Corollary 3. *The set $\mathfrak{T} \wedge \mathfrak{S}$ and hence also the set $\mathfrak{T}$ itself is not recursive* [2].

Corollary 4. *The complement of the set $\mathfrak{T}$ is not recursively enumerable* [3].

Proof. If $-\mathfrak{T}$ were recursively enumerable, it would be recursive, since the set $\mathfrak{T}$ is recursively enumerable according to theorem V **5** 1 (cf. theorem V **5** 2, p. 87).

We give still a positive formulation of theorem 1:

Theorem 5. *Let $\mathfrak{R}$ be a closed consistent class of matrices, Φ a matrix in which exactly one free variable occurs, σ a recursive definition of the function $S(x) = S(1, D_x, x)$, and Ψ the matrix $S(1, \sigma(\mathrm{a}_1), \Phi)$. Under these assumptions the integer $m = S(\sim\Psi)$ satisfies either the conditions*

$$m \in \mathfrak{R} \text{ and } \Phi(D_m) \text{ non} \in \mathfrak{R} \tag{1}$$

or the conditions

$$m \text{ non} \in \mathfrak{R} \text{ and } \sim\Phi(D_m) \text{ non} \in \mathfrak{R}. \tag{2}$$

Proof. From $m = S(\sim\Psi)$ we obtain

$$\vdash D_m \approx \sigma(D(\sim\Psi))$$

since σ is a recursive definition of S. It follows (cf. Chapter III, section **4**, theorem 1, p. 45)

$$\vdash \Phi(D_m) \leftrightarrow \Phi(\sigma(D(\sim\Psi)))$$

[1] This corollary has been first proved by Tarski [22], p. 370.

[2] This corollary is due to Rosser [19], theorem V C.

[3] Cf. Rosser [19], theorem V A.

whence by the definition of Ψ

$$\vdash \Phi(D_m) \leftrightarrow \Psi(D(\sim\Psi))$$

i.e.

$$(3) \qquad \vdash \Phi(D_m) \leftrightarrow \sim m.$$

The alternation of (1) and (2) is therefore equivalent to the trivial statement

$$(m \in \mathfrak{K})\cdot(\sim m \text{ non} \in \mathfrak{K}) \vee (m \text{ non} \in \mathfrak{K})\cdot(\sim\sim m \text{ non} \in \mathfrak{K})$$

which is obviously true since the class $\mathfrak{K}$ is consistent.

It is easy to explain why theorem 1 implies the incompleteness of the system (S): We can write down with the help of symbols of (S) matrices which from the intuitive point of view are definitions of various consistent and closed classes $\mathfrak{K}$. For instance we can take an arbitrary recursive definition of the relation

$$g \in \mathfrak{P}_{\mathfrak{A}_{\mathfrak{T}}}\cdot \tilde{g}_* = x$$

and put an existential quantifier on its front. We are then inclined to believe that we have "expressed" in (S) the arithmetical statement $x \in \mathfrak{T}$. Theorem 1 shows however that the matrix which we have thus constructed is not a recursive definition of the class $\mathfrak{T}$, and hence either there are integers in $\mathfrak{T}$ for which we cannot prove in (S) that they satisfy the matrix, or there are integers outside $\mathfrak{T}$ for which we cannot prove in (S) that they do not satisfy the matrix. In both cases we infer that there exist sentences which are intuitively true but unprovable in (S). Such sentences are certainly undecidable in (S) because provable sentences are intuitively true.

We elaborate this idea in the next section.

2. Proofs of the incompleteness theorem.

In this section we shall give three different proofs of the incompleteness theorem stating that there exist sentences undecidable in (S). A methodological discussion of these proofs will be given in section **4**.

Theorem 1. *There exist sentences Θ such that neither Θ nor $\sim\Theta$ are in $\mathfrak{T}$. Moreover a Θ with these properties can be found among*

sentences of the form $(\mathrm{Av}_k)\ \Sigma(\mathrm{v}_k)$ *where* Σ *is a recursive definition of the set of all integers.*

First proof [4]. The set $\mathfrak{S} \wedge \mathfrak{T}$ is different from $\mathfrak{S} \wedge \mathfrak{Tr}$ since the former set is definable in (S) (cf. Chapter V, section 2, theorem 10, p. 82) and the latter non-definable in (S) (cf. section 1, theorem 1, p. 88). Since $\mathfrak{T} \wedge \mathfrak{S}$ is contained in $\mathfrak{Tr} \wedge \mathfrak{S}$ (cf. Chapter IV, section 5, theorem 6, p. 69), there exist sentences Θ which belong to $\mathfrak{Tr} \wedge \mathfrak{S}$ but not to $\mathfrak{T} \wedge \mathfrak{S}$. Every such sentence is undecidable since its negation $\backsim\Theta$ does not even belong to $\mathfrak{Tr}$ (cf. Chapter IV, section 5, theorem 7, p. 69).

To find the form of the sentence Θ we remark that the set $\mathfrak{S} \wedge \mathfrak{T}$ is recursively enumerable since

$$m \in \mathfrak{S} \wedge \mathfrak{T} \equiv (\exists g)[(g \in \mathfrak{P}_{\mathfrak{A}_{\mathfrak{x}}}) \cdot (m = \bar{g}_*) \cdot (m \in \mathfrak{S})].$$

It follows that the set $\mathfrak{S} \wedge \mathfrak{T}$ possesses a formal definition of the form $(\mathrm{Ev}_k)\Phi(\mathrm{a}_1, \mathrm{v}_k)$ where Φ is a formal definition of the relation $\mathfrak{R} = \lambda mg[(g \in \mathfrak{P}_{\mathfrak{A}_{\mathfrak{x}}}) \cdot (m = \bar{g}_*) \cdot (m \in \mathfrak{S})]$ (cf. Chapter V, section 2, corollary 8, p. 80).

Using theorem 1 5 we find a Θ of the form $(\mathrm{Av}_k)\ \backsim\Phi(D_n, \mathrm{v}_k)$ such that

$$(1) \quad \begin{cases} (\Theta \in \mathfrak{Tr}) \cdot (\mathrm{Ev}_k)\Phi(D_\Theta, \mathrm{v}_k) \text{ non} \in \mathfrak{Tr} \ \vee \\ \qquad\qquad \Theta \text{ non} \in \mathfrak{Tr} \cdot \ \backsim(\mathrm{Ev}_k)\Phi(D_\Theta, \mathrm{v}_k) \text{ non} \in \mathfrak{Tr}. \end{cases}$$

The exact value of n could easily be deduced from theorem 1 5 but we shall not need it in our proof.

We maintain that the second term of the alternation (1) is false. Indeed, $\backsim(\mathrm{Ev}_k)\Phi(D_\Theta, \mathrm{v}_k)$ non $\in \mathfrak{Tr}$ implies that $(\mathrm{Ev}_k)\Phi(D_\Theta, \mathrm{v}_k)$ is in $\mathfrak{Tr}$ since the set $\mathfrak{M} \wedge \mathfrak{Tr}$ is complete (Chapter IV, section 5, theorem 9, p. 69). This means that there is a q such that $\Phi(D_\Theta, D_q) \in \mathfrak{Tr}$; Φ being a formal definition of the relation $\mathfrak{R}$, the last formula proves that $\mathfrak{R}(\Theta, q)$, i.e. that q represents a formal proof of Θ. Hence there exists a formal proof of Θ, and we obtain $\Theta \in \mathfrak{T}$ and a fortiori $\Theta \in \mathfrak{Tr}$. We have thus proved that the second

[4] This proof has been communicated to me orally by Tarski. It is contained implicite in Tarski [22], pp. 370–374.

term of the alternation (1) leads to a contradiction. It follows that the first term of this alternation is true, i.e. that

(2) $$\Theta \in \mathfrak{Tr},$$

(3) $$(\mathrm{Ev}_k)\Phi(D_\Theta, \mathrm{v}_k) \text{ non} \in \mathfrak{Tr}.$$

Observing that $\Theta = (\mathrm{Av}_k) \backsim \Phi(D_n, \mathrm{v}_k)$ we obtain from (2) $\backsim \Phi(D_n, D_q) \in \mathfrak{Tr}$ for every q. Since Φ is a recursive definition of a relation, we must have either $\vdash \Phi(D_n, D_q)$ or $\vdash \backsim \Phi(D_n, D_q)$. $\mathfrak{T}$ being a subset of $\mathfrak{Tr}$, we infer that

(4) $$\vdash \backsim \Phi(D_n, D_q) \quad \text{for } q = 1, 2, \ldots$$

Formula (4) shows that the matrix $\Sigma(\mathrm{a}_1) = \backsim \Phi(D_n, \mathrm{a}_1)$ is a recursive definition of the set of all integers. Further we infer from (3) that the sentence Θ does not belong to $\mathfrak{T}$ since otherwise there would exist a q representing a formal proof of Θ, and hence we would obtain $\vdash \Phi(D_\Theta, D_q)$ which would imply that $(\mathrm{Ev}_k)\Phi(D_\Theta, \mathrm{v}_k)$ belongs to $\mathfrak{T}$. Finally $\backsim \Theta$ does not belong to $\mathfrak{T}$ since according to (2) it is not even an element of $\mathfrak{Tr}$.

The proof of theorem 1 is thus complete.

Second proof [5]. Let Φ be a recursive definition of the relation

$$\mathfrak{R} = \lambda ng[(g \in \mathfrak{P}_{\mathfrak{A}_\mathfrak{T}}) \cdot (n = \bar{g}_*)].$$

Since $(\mathrm{Ev}_k)\Phi(\mathrm{a}_1, \mathrm{v}_k)$ is not a recursive definition of the set $\mathfrak{T}$, there exists at least one integer m such that

(5) $$\begin{cases} (m \in \mathfrak{T}) \cdot (\mathrm{Ev}_k)\Phi(D_m, \mathrm{v}_k) \text{ non} \in \mathfrak{T} \lor \\ \qquad (m \text{ non} \in \mathfrak{T}) \cdot \backsim (\mathrm{Ev}_k)\Phi(D_m, \mathrm{v}_k) \text{ non} \in \mathfrak{T}. \end{cases}$$

The exact value of m is irrelevant for the proof, but we may remark for the sake of completeness that $m = S(1, D(\Psi), \Psi)$ where $\Psi = (\mathrm{Av}_k) \backsim \Phi(\sigma(\mathrm{a}_1), \mathrm{v}_k)$ and where σ is a recursive definition of the function $S(1, D_x, x)$ (cf. section 1, theorem 5, p. 89).

We shall show that the first term of the alternation (5) is false. Indeed, if m were an element of $\mathfrak{T}$, there would be an integer q such that $\mathfrak{R}(m, q)$ which proves that $\Phi(D_m, D_q)$ would belong to $\mathfrak{T}$,

[5] This is approximatively the original proof of Gödel [9], pp. 187–189.

and hence $(\mathrm{Ev}_k)\Phi(D_m, \mathrm{v}_k)$ would be an element of $\mathfrak{T}$. This shows that the first term of (5) leads to a contradiction.

It follows that

$$(6) \qquad m \text{ non } \in \mathfrak{T},$$

$$(7) \qquad (\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k) \text{ non } \in \mathfrak{T}.$$

For every q the sentence $\backsim\Phi(D_m, D_q)$ is an element of $\mathfrak{T}$. Indeed, if $\backsim\Phi(D_m, D_q)$ were not in $\mathfrak{T}$, then $\Phi(D_m, D_q)$ would be in $\mathfrak{T}$ since Φ is a recursive definition of a relation. Hence we would obtain $\mathfrak{R}(m, q)$ which would imply that $m \in \mathfrak{T}$. Since this contradicts (6) we infer that

$$(8) \qquad \backsim\Phi(D_m, D_q) \in \mathfrak{T} \text{ for } q = 1, 2, \ldots$$

Formula (8) proves that the sentence $\Theta = (\mathrm{Ev}_k)\Phi(D_m, \mathrm{v}_k)$ does not belong to $\mathfrak{T}$; indeed if this sentence were in $\mathfrak{T}$, the set $\mathfrak{T}$ would be ω-inconsistent. Formula (7) shows on the other hand that $\backsim\Theta$ does not belong to $\mathfrak{T}$. Finally, it follows from (8) that the matrix $\Sigma(\mathrm{a}_1) = \backsim\Phi(D_m, \mathrm{a}_1)$ is a recursive definition of the set of all integers. Since $\vdash \backsim\Theta \leftrightarrow (\mathrm{Av}_k)\Sigma(\mathrm{v}_k)$, we see that the sentence $\backsim\Theta$ satisfies all the conditions which we required in theorem 1.

Third proof [6]. Let us consider the same recursive relation $\mathfrak{R}$ as in the second proof, and let Π be a recursive definition of $\mathfrak{R}$ and ν a recursive definition of the function $\backsim x$. It follows from these definitions that

$$(9) \qquad m \in \mathfrak{T} \equiv (\exists g)\mathfrak{R}(m, g),$$

$$(10) \qquad \mathfrak{R}(m, g) \supset \vdash \Pi(D_m, D_g),$$

$$(11) \qquad \sim \mathfrak{R}(m, g) \supset \vdash \backsim\Pi(D_m, D_g),$$

$$(12) \qquad \vdash \nu(D(m)) \approx D(\backsim m).$$

Let v_h be a bound variable which does not occur in Π, and put

$$(13) \qquad \mathrm{P} = \Pi(\mathrm{a}_1, \mathrm{a}_2) \;\&\; (\mathrm{Av}_h)[\mathrm{v}_h \lessdot \mathrm{a}_2 \rightarrow \backsim\Pi(\nu(\mathrm{a}_1), \mathrm{v}_h)].$$

We maintain that P is a recursive definition of $\mathfrak{R}$. Indeed, if

[6] This proof is due to Rosser [19], theorem II, p. 89.

$\sim \mathfrak{R}(m, g)$, then $\vdash \backsim \Pi(D_m, D_g)$ according to (11) and since $\vdash \backsim \Pi \rightarrow \backsim \mathrm{P}$, it follows that $\vdash \backsim \mathrm{P}(D_m, D_g)$.

Let us now assume that $\mathfrak{R}(m, g)$. Since $\mathfrak{T}$ is a consistent class (cf. Chapter IV, section **5**, theorem 8, p. 69) it follows that no i satisfies the condition $\mathfrak{R}(\backsim m, i)$ and hence we obtain by (10) and (11)

(14) $\vdash \Pi(D_m, D_g)$,

(15) $\vdash \backsim \Pi(D(\backsim m), D_i)$ for $i = 1, 2, 3, \ldots.$

Formulas (15) and (12) together with theorem III **4** 1 yield

(16) $\vdash \backsim \Pi(\nu(D_m), D_i)$ for $i = 1, 2, 3, \ldots,$

and we obtain by the propositional calculus

$$\vdash [\mathrm{a}_3 \approx D_1 \vee \mathrm{a}_3 \approx D_2 \vee \ldots \vee \mathrm{a}_3 \approx D_{g-1}] \rightarrow \backsim \Pi(\nu(D_m), \mathrm{a}_3)$$

which implies (according to theorem III **7** 7) that

(17) $(\mathrm{Av}_h)[\mathrm{v}_h \lhd D_g \rightarrow \backsim \Pi(\nu(D_m), \mathrm{v}_h)].$

The conjunction of (14) and (17) gives according to (13) $\vdash \mathrm{P}(D_m, D_g)$. Hence P is a recursive definition of the relation $\mathfrak{R}$.

We now apply theorem **1** 5 taking $\mathfrak{K} = \mathfrak{T}$ and

(18) $\Phi = (\mathrm{Ev}_k)\mathrm{P}(\mathrm{a}_1, \mathrm{v}_k)$

where v_k is a bound variable which does not occur in P. We obtain a sentence Θ such that either

(19) $\Theta \in \mathfrak{T}$ and $\Phi(D_\Theta)$ non $\in \mathfrak{T}$

or

(20) Θ non $\in \mathfrak{T}$ and $\backsim\Phi(D_\Theta)$ non $\in \mathfrak{T}$.

Moreover, (cf. formula (3) in the proof of theorem **1** 5)

(21) $\vdash \backsim \Theta \leftrightarrow \Phi(D_\Theta)$.

From the two possibilities (19) and (20) the first leads to a contradiction. Indeed, from $\Theta \in \mathfrak{T}$ it follows that $\mathfrak{R}(\Theta, g)$ for some g whence, by the property of P which we have just established,

$\vdash \mathrm{P}(D_\Theta, D_g)$ and hence (cf. Chapter III, section 3, theorem 1, p. 44) $\vdash (\mathrm{E}\mathrm{v}_k)\mathrm{P}(D_\Theta, \mathrm{v}_k)$ i.e. $\Phi(D_\Theta) \in \mathfrak{T}$. Since this contradicts the second part of (19), we see that (19) is impossible and therefore that (20) must hold.

The same reasoning shows also that

$$(22) \qquad \vdash \backsim \mathrm{P}(D_\Theta, D_g) \quad \text{for } g = 1, 2, 3, \ldots .$$

We can now prove that Θ is an undecidable sentence. Since Θ non $\in \mathfrak{T}$ by (20), it remains to show that $\backsim \Theta$ non $\in \mathfrak{T}$.

Assume that

$$(23) \qquad \backsim \Theta \in \mathfrak{T}$$

and let g be an integer such that $\mathfrak{R}(\backsim \Theta, g)$. We obtain therefore

$$(24) \qquad \vdash \mathrm{P}(D(\backsim \Theta), D_g).$$

By (21) and (23) we obtain $\vdash \Phi(D_\Theta)$ whence by (18) and theorem III 7 11

$$\vdash (\mathrm{E}\mathrm{v}_k)\{[\mathrm{v}_k \lessdot D_g \,\&\, \mathrm{P}(D_\Theta, \mathrm{v}_k)] \lor [\mathrm{v}_k \approx D_g \,\&\, \mathrm{P}(D_\Theta, \mathrm{v}_k)] \lor [D_g \lessdot \mathrm{v}_k \,\&\, \mathrm{P}(D_\Theta, \mathrm{v}_k)]\}.$$

Using the law of distributivity of the existential quantifier over the alternation we obtain from the last formula

$$(25) \qquad \begin{cases} \vdash (\mathrm{E}\mathrm{v}_k)[\mathrm{v}_k \lessdot D_g \,\&\, \mathrm{P}(D_\Theta, \mathrm{v}_k)] \lor \\ \quad (\mathrm{E}\mathrm{v}_k)[\mathrm{v}_k \approx D_g \,\&\, \mathrm{P}(D_\Theta, \mathrm{v}_k)] \lor \\ \quad (\mathrm{E}\mathrm{v}_k)[D_g \lessdot \mathrm{v}_k \,\&\, \mathrm{P}(D_\Theta, \mathrm{v}_k)]. \end{cases}$$

Let us write this formula as $\Gamma_1 \lor \Gamma_2 \lor \Gamma_3$. We shall show that $\vdash \backsim \Gamma_3$. Indeed it follows from (13) that

$$\vdash \mathrm{P}(D_\Theta, \mathrm{a}_2) \,\&\, D_g \lessdot \mathrm{a}_2 \to \backsim \Pi(D(\backsim \Theta), D_g)$$

and from (24) that $\vdash \Pi(D(\backsim \Theta), D_g)$. Using the law of transposition we get $\vdash \backsim [\mathrm{P}(D_\Theta, \mathrm{a}_2) \,\&\, D_g \lessdot \mathrm{a}_2]$ whence

$$\vdash (\mathrm{A}\mathrm{v}_k) \backsim [\mathrm{P}(D_\Theta, \mathrm{v}_k) \,\&\, D_g \lessdot \mathrm{v}_k] \text{ i.e. } \vdash \backsim \Gamma_3.$$

It is known from the propositional calculus that if $\vdash \Gamma_1 \lor \Gamma_2 \lor \Gamma_3$

and $\vdash \backsim \Gamma_3$, then $\vdash \Gamma_1 \vee \Gamma_2$. Hence the last term in (25) can be omitted. Using theorems III 4 1 and III 7 7 we obtain now from (25) the formula

$$\vdash \mathrm{P}(D_\Theta, D_1) \vee \mathrm{P}(D_\Theta, D_2) \vee \ldots \vee \mathrm{P}(D_\Theta, D_g).$$

From (22) we obtain however

$$\vdash \backsim \mathrm{P}(D_\Theta, D_1) \ \& \ \backsim \mathrm{P}(D_\Theta, D_2) \ \& \ \ldots \ \& \ \backsim \mathrm{P}(D_\Theta, D_g)$$

Hence the assumption (23) leads to the result that the class $\mathfrak{T}$ is inconsistent. This assumption is therefore wrong and the sentence Θ undecidable.

The sentence Θ has the form $(\mathrm{Av}_k)\Sigma(\mathrm{v}_k)$ where the matrix $\Sigma(\mathrm{a}_1) = \backsim \mathrm{P}(D_\Theta, \mathrm{a}_1)$ is according to (22) a recursive definition of the set of all positive integers.

The proof of theorem 1 is thus complete.

We note still separately for later use the result established in the second proof. We divide this result into two parts one of which uses only the consistency and the other the ω-consistency of $\mathfrak{T}$:

Theorem 2. *If the set $\mathfrak{T}$ is consistent, Φ is a recursive definition of the relation $\mathfrak{R}$ (or more generally of any recursive relation $\mathfrak{P}$ such that $m \in \mathfrak{T} \equiv (\exists n)\mathfrak{P}(m, n)$), σ a recursive definition of the function $S(x) = S(1, D_x, x)$ and $\Psi = (\mathrm{Av}_k) \backsim \Phi(\sigma(\mathrm{a}_1), \mathrm{v}_k)$, $m = S(1, D(\Psi), \Psi)$, then*

$$(26) \qquad (\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k) \text{ non} \in \mathfrak{T},$$

$$(27) \qquad \backsim \Phi(D_m, D_q) \in \mathfrak{T} \text{ for } q = 1, 2, 3, \ldots$$

$$(28) \qquad (\mathrm{Av}_k) \backsim \Phi(\sigma(D(\Psi)), \mathrm{v}_k) \leftrightarrow (\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k) \in \mathfrak{T}.$$

Theorem 3. *If the set $\mathfrak{T}$ is ω-consistent, then the sentence $(\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k)$ is undecidable.*

Note that the formula (28) is the consequence of the definition of m and the formula $\vdash D_m \approx \sigma(D(\Psi))$ which results immediately from the definition of σ.

Note further that

$$(29) \quad \left\{ \begin{aligned} (\mathrm{Av}_k) \backsim \Phi(\sigma(D(\Psi)), \mathrm{v}_k) &= S(1, D(\Psi), (\mathrm{Av}_k) \backsim \Phi(\sigma(\mathrm{a}_1), \mathrm{v}_k)) \\ &= S(1, D(\Psi), \Psi) = m. \end{aligned} \right.$$

3. Generalizations. The three proofs given in section 2 are evidently very akin to each other and use one and the same fundamental idea. However, each of them uses different properties of the set $\mathfrak{T}$ and the relation $\mathfrak{R}$. It follows that the three proofs will give rise to three distinct theorems when we enumerate separately those assumptions concerning the set $\mathfrak{T}$ which are actually needed to carry out the proof. In this way we obtain the following theorems:

Theorem 1. *No definable closed subclass* $\mathfrak{K}$ *of* $\mathfrak{Tr} \wedge \mathfrak{M}$ *is complete* [7].

To prove this theorem we proceed exactly as in the first part of the first proof in section 2.

The form of the sentence undecidable with respect to $\mathfrak{K}$ depends on the class $\mathfrak{K}$, and more exactly, on the form of matrices which can be taken as formal definitions of $\mathfrak{K}$. In the special case, however, when the class $\mathfrak{K}$ is recursively enumerable, i.e. has the form $\lambda n[(\exists q)\mathfrak{Q}(n, q)]$ with recursive $\mathfrak{Q}$, the proof given on pp. 91—92 applies without changes with $\mathfrak{R}$ replaced by $\mathfrak{Q}$ and we obtain a sentence undecidable with respect to $\mathfrak{K}$ of the form $(\mathrm{Av}_k)\Sigma(\mathrm{v}_k)$ where Σ is a recursive definition of the set of all positive integers.

Theorem 2. *No recursively enumerable, closed and consistent class of matrices is complete* [8].

Proof. A recursively enumerable class has the form

$$\lambda n[(\exists q)\mathfrak{Q}(n, q)]$$

where $\mathfrak{Q}$ is a recursive relation. It is now sufficient to repeat the third proof of section 2 replacing everywhere $\mathfrak{R}$ by $\mathfrak{Q}$.

The undecidable sentence which we obtain in this way has exactly the same form as the sentence obtained in theorem 2 1.

Slightly less obvious is the generalization of the second proof.

Theorem 3. *No definable closed and ω-consistent class $\mathfrak{K}$ is complete.*

Proof. We first remark that there exist definable sets which

[7] This general formulation of incompleteness theorems is due to Tarski [22], p. 370 and [23], p. 109.

[8] This theorem is due to Rosser [19], theorem II, p. 89.

are not $\mathfrak{K}$-definable (e.g. the set $\mathfrak{K}$ itself, cf. section 1, theorem 1, p. 88). Using theorem V 4 1 we infer that there exists a $\mathfrak{K}$-definable relation $\mathfrak{R}$ with, say, n arguments such that for some $j \leqslant n$ the function $F = \min_j \mathfrak{R}$ is not $\mathfrak{K}$-definable.

Let Φ be a $\mathfrak{K}$-definition of $\mathfrak{R}$, v_h a *bound variable* which does not *occur* in Φ and $\varphi = (\mu \mathrm{v}_h) S(j, \mathrm{v}_h, \Phi)$. Since φ is not a $\mathfrak{K}$-definition of F, there exist integers $p_1, p_2, \ldots, p_{n-1}$ such that

$$(1) \qquad \varphi(D(p_1), \ldots, D(p_{n-1})) \approx D(F(p_1, \ldots, p_{n-1})) \text{ non } \in \mathfrak{K}.$$

We maintain that $F(p_1, \ldots, p_{n-1}) = 1$ and that

$$(2) \qquad \sim \mathfrak{R}(p_1, \ldots, p_{j-1}, q, p_j, \ldots, p_{n-1}) \text{ for } q = 1, 2, \ldots.$$

Indeed, if there were a smallest q for which (2) would be false, we would obtain $F(p_1, \ldots, p_{n-1}) = q$ and the following relations would hold:

$$\Phi(D(p_1), \ldots, D(p_{j-1}), D_q, D(p_j), \ldots, D(p_{n-1})) \in \mathfrak{K},$$
$$\backsim \Phi(D(p_1), \ldots, D(p_{j-1}), D_r, D(p_j), \ldots, D(p_{n-1})) \in \mathfrak{K} \; (r < q)$$

which proves (cf. Chapter III, section 7, theorem 12, p. 53) that $\varphi(D(p_1), \ldots, D(p_{n-1})) \approx D_q \in \mathfrak{K}$. Since this contradicts (1), we obtain (2).

Let us put

$$\bar{\Phi} = \Phi(D(p_1), \ldots, D(p_{j-1}), \mathrm{a}_j, D(p_j), \ldots, D(p_{n-1}))$$

and

$$\bar{\varphi} = (\mu \mathrm{v}_h) S(j, \mathrm{v}_h, \bar{\Phi}).$$

From (2) we obtain

$$\backsim \bar{\Phi}(D_q) \in \mathfrak{K} \text{ for } q = 1, 2, 3, \ldots$$

and hence

$$(3) \qquad (\mathrm{E}\mathrm{v}_h) S(j, \mathrm{v}_h, \bar{\Phi}) \text{ non } \in \mathfrak{K}$$

since the set $\mathfrak{K}$ is ω-consistent. Using theorem III 3 3, we obtain therefore $\backsim \bar{\varphi} \approx 1 \text{ non } \in \mathfrak{K}$. Since (1) can be written as $\bar{\varphi} \approx 1 \text{ non } \in \mathfrak{K}$ we infer that the *sentence* $\bar{\varphi} \approx 1$ is undecidable with respect to $\mathfrak{K}$.

Theorem 3 is thus proved.

Remark. It is easy to show that also the sentence $\Theta = (\mathrm{Ev}_h)S(j, \mathrm{v}_h, \overline{\Phi})$ satisfies the conditions Θ non $\in \mathfrak{K}$, $\backsim\Theta$ non $\in \mathfrak{K}$. First formula results immediately from (3). If the second were false, we would obtain $(\mathrm{Av}_h)S(j, \mathrm{v}_h, \backsim\overline{\Phi}) \in \mathfrak{K}$ which would imply (cf. Chapter III, section 3, theorem 3, p. 45) that $\overline{\varphi} \approx 1 \in \mathfrak{K}$. Since this contradicts our previous result, we infer that $\backsim\Theta$ non $\in \mathfrak{K}$.

4. Discussion of the incompleteness theorems. In this section we compare the three theorems obtained in section 3.

First we show that theorems 3 2 and 3 3 are not comparable. To achieve this we must compare the hypotheses concerning $\mathfrak{K}$ under which theorems 3 2 and 3 3 have been proved. We see that one hypothesis of theorem 3 2 (recursive enumerability of $\mathfrak{K}$) is stronger than the corresponding hypothesis of theorem 3 3 (definability of $\mathfrak{K}$). The other pair of hypotheses behaves conversely: The assumption of consistency made in theorem 3 2 is weaker than the assumption of ω-consistency made in theorem 3 3.

In order to make this heuristic argument really convincing we have to prove the following two theorems:

Theorem 1. *There exists a definable and ω-consistent closed class $\mathfrak{K}_1$ of matrices which is not recursively enumerable.*

Theorem 2. *There exists a recursively enumerable and consistent closed class $\mathfrak{K}_2$ of matrices which is not ω-consistent.*

Proof of theorem 1. Let $\mathfrak{D}$ be the set of all recursive definitions of the set of positive integers, i.e.

$$(1)\qquad \begin{cases} \Phi \in \mathfrak{D} \equiv (\Phi \in \mathfrak{M}) \cdot (n)[Oc(\mathrm{a}_n, \Phi) = 2 \equiv n = 1] \cdot \\ \qquad\qquad (n)[S(1, D_n, \Phi) \in \mathfrak{T}]. \end{cases}$$

Let $\mathfrak{L}$ consist of all sentences of the form $(\mathrm{Av}_k)\Sigma(\mathrm{v}_k)$ where $\Sigma \in \mathfrak{D}$ and v_k is a bound variable which does not occur in Σ. Finally put $\mathfrak{K}_1 = \mathfrak{T}_{\mathfrak{L}}$.

The set $\mathfrak{L}$ is contained in $\mathfrak{Tr}$ since for every Σ in $\mathfrak{D}$ the sentences $S(1, D_n, \Sigma)$ $(n = 1, 2, \ldots)$ are all provable and hence $(\mathrm{Av}_k)\Sigma(\mathrm{v}_k) \in \mathfrak{Tr}$ (cf. Chapter IV, section 5, theorem 12, p. 72). It follows that

the set $\mathfrak{K}_1$ is contained in $\mathfrak{Tr}$ and hence ω-consistent (Chapter IV, section 5, theorem 11, p. 71).

The definability of $\mathfrak{L}$ is an immediate consequence of the formula (1) and the equivalence

$$\Omega \in \mathfrak{L} \equiv (\exists \Sigma)(\exists k)\{(\Sigma \in \mathfrak{D})\cdot[Oc(\mathrm{v}_k, \Sigma) = 1]\cdot [\Omega = (\mathrm{Av}_k)S(1, \mathrm{v}_k, \Sigma)]\}.$$

It follows that also the class $\mathfrak{K}_1$ is definable (cf. Chapter V, section 3, theorem 10, p. 82).

If $\mathfrak{K}_1$ were a recursively enumerable class, we could apply theorem 3 2 to it and would obtain a sentence $(\mathrm{Av}_k)\Sigma(\mathrm{v}_k)$ undecidable with respect to $\mathfrak{K}_1$ and such that $\Sigma \in \mathfrak{D}$. This, however, is not possible since all sentences of this form are $\mathfrak{K}_1$-provable. Hence the class $\mathfrak{K}_1$ is not recursively enumerable and theorem 1 is proved.

Proof of theorem 2. Let Σ be a recursive definition of the set of all integers such that the sentence $(\mathrm{Av}_k)\Sigma(\mathrm{v}_k)$ be undecidable (v_k is here an arbitrary bound variable which does not occur in Σ). Put

$$\Phi = S(1, (\mu \mathrm{v}_k)S(1, \mathrm{v}_k, \backsim \Sigma), \backsim \Sigma) = (\mathrm{Ev}_k) \backsim \Sigma(\mathrm{v}_k).$$

Further, let $\mathfrak{L}$ be the unit class consisting of Φ alone and put $\mathfrak{K}_2 = \mathfrak{T}_{\mathfrak{L}}$.

Since the set $\mathfrak{L}$ is recursive, the class $\mathfrak{K}_2$ is recursively enumerable (cf. Chapter V, section 5, theorem 1, p. 86).

$\mathfrak{K}_2$ is ω-inconsistent. Indeed, the sentences $\Sigma(D_1)$, $\Sigma(D_2)$, ..., $\Sigma(D_n)$, ... are all provable and hence belong to $\mathfrak{K}_2$. If $\mathfrak{K}_2$ were ω-consistent, the sentence $(\mathrm{Ev}_k) \backsim \Sigma(\mathrm{v}_k)$ would not belong to $\mathfrak{K}_2$ which would lead to a contradiction since this sentence is identical with Φ.

Finally $\mathfrak{K}_2$ is consistent, since the sentence Φ is undecidable (cf. Chapter II, section 6, theorem 1, p. 41).

Theorem 2 is thus proved. We note still that $\mathfrak{K}_2$ contains false sentences, e.g. Φ.

A similar discussion shows also that theorems 3 1 and 3 2 are not comparable. We have again in these two theorems pairs of

conditions imposed on a closed class $\mathfrak{K}$ of matrices. We see that one condition imposed on $\mathfrak{K}$ in theorem 3 1 ($\mathfrak{K} \subset \mathfrak{Tr}$) is stronger than the corresponding condition of theorem 3 2 (consistency of $\mathfrak{K}$), whereas the second pair of hypotheses behaves conversely. This heuristic argument is supported by the following theorems:

Theorem 3. *There exists a definable closed subclass of $\mathfrak{Tr}$ which is not recursively enumerable.*

Theorem 4. *There exists a recursively enumerable closed and consistent class which is not contained in $\mathfrak{Tr}$.*

Proof. The class $\mathfrak{K}_1$ satisfies the conditions of theorem 3 and the class $\mathfrak{K}_2$ the conditions of theorem 4.

It remains to compare theorems 3 1 and 3 3. We shall show that theorem 3 3 is stronger than theorem 3 1.

First of all it is evident that theorem 3 3 is at least as strong as theorem 3 1 since subsets of $\mathfrak{Tr}$ are always ω-consistent.

To show that the converse implication does not hold we shall prove the following theorem:

Theorem 5. *There exists a closed class $\mathfrak{K}$ of matrices which is ω-consistent but not contained in $\mathfrak{Tr}$.*

Proof. Let us denote by $\{\Phi\}$ the class whose sole element is Φ. We shall call a sentence Φ ω-consistent if the set $\mathfrak{T}_{\{\Phi\}}$ is ω-consistent.

The following equivalence results immediately from the definition of ω-consistency:

$$[\Phi \textit{ is } \omega\textit{-consistent}] \equiv \Phi \in \mathfrak{S}\cdot(\Psi)\{(\Psi \in \mathfrak{M})\cdot$$
$$(h)\ [Oc(\mathrm{a}_h,\Psi) = 2 \equiv h = 1]\cdot(n)[S(1, D_n,\Psi) \in \mathfrak{T}_{\{\Phi\}}] \rightarrow$$
$$(k)\{[Oc(\mathrm{v}_k,\Psi) = 1] \rightarrow (\mathrm{Ev}_k)[S(1, \mathrm{v}_k, \backsim\Psi) \text{ non} \in \mathfrak{T}_{\{\Phi\}}]\}\}.$$

Since $n \in \mathfrak{T}_{\{\Phi\}} \equiv (\exists g)[(g \in \mathfrak{P}_{\{\Phi\}})\cdot(\bar{g}_* = n)]$ and the class $\mathfrak{P}_{\{\Phi\}}$ is recursive (cf. Chapter V, section 3, theorem 9, p. 82) this equivalence proves that the set $\mathfrak{C}$ of ω-consistent sentences is definable. Since the set $\mathfrak{S} \wedge \mathfrak{Tr}$ is not definable according to theorem 1 1, it follows that the sets $\mathfrak{C}$ and $\mathfrak{S} \wedge \mathfrak{Tr}$ do not coincide, and since $\mathfrak{S} \wedge \mathfrak{Tr}$ is a subset of $\mathfrak{C}$, there exist sentences which are ω-con-

sistent but not true. For every such sentence Φ the set $\mathfrak{T}_{\{\Phi\}}$ is ω-consistent but not contained in $\mathfrak{Tr}$.

Theorem 5 is thus proved.

Remark. Using theorem 1 5 one can easily exhibit explicitly an ω-consistent and false sentence Φ. It can be shown that the Φ obtained in this way has the form $(\mathrm{Ev}_m)(\mathrm{Av}_n)(\mathrm{Ev}_p)\Omega(\mathrm{v}_m, \mathrm{v}_n, \mathrm{v}_p)$ where Ω is a recursive definition of a ternary relation. It is worth while to note in connection with this result that every ω-consistent sentence of the form $(\mathrm{Av}_n)(\mathrm{Ev}_p)\Omega(\mathrm{v}_n, \mathrm{v}_p)$ where Ω is a recursive definition of a binary relation is automatically true [9].

We remark still that results reached in theorems 3 2 and 3 3 are in a sense best possible. It is namely known that there exist definable, consistent, and complete classes of matrices [10]. Hence we can neither replace in theorem 3 2 the assumption of recursive enumerability of $\mathfrak{R}$ by the weaker assumption of definability nor in theorem 3 3 the assumption of ω-consistency of $\mathfrak{R}$ by the weaker assumption of the ordinary consistency.

[9] Turing [25], p. 194.

[10] The existence of such classes of matrices has been first established by Lindenbaum. Cf. Tarski [20], theorem I. 56, p. 394. The proof that classes constructed by Lindenbaum are definable is contained in the paper Novak [15], p. 95–99.

APPENDIX

SOME FURTHER RESULTS OF GÖDEL'S THEORY

The aim of this appendix is to review some important theorems which are closely related to the theory developed in the main text but which require a rather elaborate technique and for that reason could not be included into our systematic exposition.

We shall not give full proofs of theorems in question but content ourselves with a sketch of the main ideas of proofs. The author hopes that these sketches are detailed enough to convince the reader of the truth of the theorems.

1. Undecidability of the sentence expressing the consistency of (S). The theorem which will be dealt with in the present section has its source in the following investigation.

All theorems established in the previous Chapters have purely arithmetical character and their proofs do not use any non-arithmetical notion. In particular we know from Chapter II, pp. 27—29 that the "linguistic" terms such as s e n t e n c e, p r o o f, s u b s t i t u t i o n etc., printed in spaced characters, are names of certain arithmetical functions, classes or relations and that their definitions are expressible in purely arithmetical terms.

Had we decided to write down all our theorems without abbreviations printed in spaced characters, it would become obvious that most of the theorems proved in Chapters I—VI are expressible by means of the logical connectives, quantifiers, and equations between arithmetical functions definable in terms of addition and multiplication alone. Only theorems which belong to semantics require still other notions namely that of an arbitrary set of integers as well as the membership-relation.

We shall at present disregard the semantical part of our theory and consider only theorems which are expressible in purely arithmetical terms. It is usual to call these theorems *syntactical*.

To every syntactical theorem we let correspond a sentence of (S) which we shall call its arithmetical counterpart. The rule of forming this sentence can be described as follows.

Let T be a syntactical theorem. We let correspond to every variable occurring in T a bound variable in such a way that to different variables correspond different elements of $\mathfrak{Vb}$. To every expression of the form $a + b = c$ or $a \cdot b = c$ which is a part of T we let correspond the matrix-forms

$$v_k \mathbin{\boldsymbol{+}} v_l \approx v_m \text{ or } v_k \times v_l \approx v_m$$

where the bound variables v_k, v_l, v_m correspond to "a", "b", "c". Finally, we agree that if matrix-forms Φ_1, Φ_2 correspond to expressions E_1, E_2, then $\Phi_1 \rightarrow \Phi_2$ corresponds to $E_1 \supset E_2$, $\backsim \Phi_1$ corresponds to $\sim E_1$ and $(Av_k)\Phi_1$ corresponds to $(n)E_1$ where v_k is the bound variable which corresponds to "n" [1].

It is now natural to ask what sentences will be obtained as the arithmetical counterparts of the various syntactical theorems and whether or not these sentences are provable in (S).

We limit our discussion to theorem VI 2 2 or rather to a particular case of this theorem which we obtain choosing for Φ a particular recursive definition of the relation $\mathfrak{R}$. Meanwhile we leave Φ arbitrary.

The arithmetical counterpart of theorem VI 2 2 has the form of an implication

$$(1) \qquad \Delta \rightarrow \Gamma$$

[1] In order to make this definition precise it would be necessary to present the intuitive arithmetic in which the syntactical theorems are expressed and proved in the form of a formal system. We remark however that we are not interested in the general theory of correspondence between the syntactical theorems and sentences of (S) but wish only to write down the arithmetical counterparts of the syntactical theorems of Chapters I—VI. Since these theorems are finite in number, we can with a little patience write down effectively their arithmetical counterparts. The definition given above is to be understood as an instruction how this has to be done. Once these arithmetical counterparts (which are sentences of (S), and hence certain well defined integers) have been calculated we can wholly forget the definition of correspondence.

where Δ is the arithmetical counterpart of the statement

(2) *the set $\mathfrak{T}$ is consistent*

and Γ is the arithmetical counterpart of the statement

(3) *the sentence* $(\mathrm{Av}_k) \backsim \Phi(\sigma(D_\Psi), \mathrm{v}_k)$ *is not in* $\mathfrak{T}$.

It is easy to convince oneself that Δ has the form

$$\Delta = (\mathrm{Av}_1) \backsim \Phi_0(D(\backsim 1 \approx 1), \mathrm{v}_1)$$

where Φ_0 is a recursive definition of the relation $\mathfrak{R}$. We fix now definitely the matrix Φ taking $\Phi = \Phi_0$, and ask what will be the arithmetical counterpart of the syntactical statement (3).

First of all we can assert that this counterpart is

$$\Gamma = (\mathrm{Av}_k) \backsim \Phi_0(D[(\mathrm{Av}_k) \backsim \Phi_0(\sigma(D(\Psi)), \mathrm{v}_k)], \mathrm{v}_k).$$

Indeed, Γ has to express in the language of (S) the non-existence of a formal proof of the sentence $\Sigma = (\mathrm{Av}_k) \backsim \Phi_0(\sigma(D(\Psi)), \mathrm{v}_k)$, and hence it must have the form $(\mathrm{Av}_k) \backsim \Phi_0(D(\Sigma), \mathrm{v}_k)$.

In VI 2 (29) we have shown that

$$(\mathrm{Av}_k) \backsim \Phi_0(\sigma(D(\Psi)), \mathrm{v}_k) = m.$$

It follows that

$$\Gamma = (\mathrm{Av}_k) \backsim \Phi_0(D(m), \mathrm{v}_k),$$

i.e. that Γ is the very same sentence which was shown unprovable in (S) provided that this system is consistent [2].

Having thus calculated the arithmetical counterpart of theorem VI 2 2, we ask now what are the axioms and rules of proof which are necessary to prove this sentence.

In the proof of theorem VI 2 2 we have used freely the axioms and theorems of intuitive arithmetic. All these axioms and theorems, however, have arithmetical counterparts in (S), and these counterparts are axioms or theorems of (S). In other words, the proof of theorem VI 2 2 which we have carried out in

[2] For that reason one often says that the undecidable sentence constructed in theorem VI 2 2 "asserts its own unprovability".

the intuitive arithmetic can be repeated word by word in the system (S) and yields a formal proof of the sentence (1).

Summing up this discussion we obtain

Theorem 1. $\vdash \Delta \rightarrow \Gamma$.

Our proof of this important theorem is evidently very imperfect. To make it complete we would have to exhibit explicitly a formal proof of the sentence (1); in other words we would have to present the whole theory developed in Chapters I, II, III, V, and VI up to the proof of theorem VI 2 2 inclusive in the formal language of the system (S). Such an enterprise would require much space but is not difficult in itself. At any rate the method of proof is sufficiently well elucidated by the remarks given above.

From theorem 1 we easily obtain

Theorem 2. *If the set $\mathfrak{T}$ is consistent, the sentence Δ is not provable* [3].

Proof. From $\vdash \Delta$ follows by theorem 1 $\vdash \Gamma$ and since this implies the inconsistency of $\mathfrak{T}$ according to theorem VI 2 2, we infer that non $\vdash \Delta$.

To explain the point which constitutes the importance of theorem 2 we make the following remarks.

According to Hilbert the most important task of theoretical logic is the search for a finitary proof that formalized arithmetic is consistent. Neither Hilbert nor any member of his school has ever explained precisely what is meant by the word "finitary". One could suspect, not without reasons, that if a proof is finitary, it has an arithmetical counterpart in the system (S). If we would agree on this definition, then we could infer from theorem 2 that no "finitary" consistency proof for arithmetic exists. Indeed such a proof being formalizable within (S) would yield a formal proof of the sentence Δ which would imply that (S) is inconsistent.

Our argument shows that either the word "finitary" has been wrongly defined or Hilbert's program is impossible.

We do not intend to discuss here other possible definitions of

[3] Gödel [9], p. 196–197, Satz XI.

"finitary proofs" [4]. All what we wanted to show is that Gödel's theory establishes, so to say, a "lower limit" for the meaning of the word "finitary" for which the Hilbert's program is still possible.

It is interesting to note that we have found in Chapter IV a consistency proof for arithmetic (cf. Chapter IV, section **5**, theorem 8, p. 69) but that this proof used semantical notions which, as we remarked above, do not possess arithmetical counterparts in (S).

2. Problem of truth of undecidable sentences [5]. We shall now take up the problem of intuitive truth of the undecidable sentences constructed in Chapter VI, and, in particular, of the sentence obtained in theorem VI 2 2.

The intuitive content of the sentence $(\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k)$ is that there exists no formal proof for the sentence m. Since m non $\in \mathfrak{T}$ according to theorem VI **2** 2, we see that the sentence $(\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k)$ is intuitively correct.

The same result can be deduced from formula (27) of theorem VI **2** 2 which yields according to theorem IV **5** 12 the following

Theorem 1. $(\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k) \in \mathfrak{Tr}$.

We see that the sentence $(\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k)$ is intuitively obvious and does not represent any difficult mathematical problem the solution of which would surpass our mathematical knowledge. The undecidability of this sentence has its roots not in the difficulty of the problem which it represents but in the insufficiency of the methods of proof admitted in (S). In fact, the undecidability of this sentence does not show anything more than that the system (S) is not an adequate formalization of the intuitive arithmetic.

These considerations suggest that we should seek for new rules of proof which together with the rules already admitted in (S) would guarantee the decidability of the sentence $(\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k)$.

[4] It follows from certain results of Gentzen [8] that there are reasonable meanings of the word "finitary" such that a finitary consistency proof of arithmetic is indeed possible.

[5] Results presented in this section are due to Tarski [22], pp. 400–403. Cf. also footnote [48a]) on p. 191 of Gödel [9].

Such additional rules of inference can be obtained in many different ways. The simplest method would be to adjoin the sentence $(\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k)$ or the sentence Δ discussed in section 1 to the set of axioms. Much more interesting however is another method which uses the semantical theory of Chapter IV.

As we have noted in section 1 the semantical notions do not have arithmetical counterparts in (S) since they operate with the general notion of set. However, we can construct easily other formal systems of arithmetic in which semantics of (S) is formalizable just as the syntax of (S) is formalizable within (S).

We shall outline here one such formal system.

First we change the definitions of Chapter II 2, p. 27 putting

$$\mathrm{v}_h = 8h, \quad \mathrm{a}_h = 8h + 1, \quad h = 1, 2, 3, \ldots,$$

and replacing in the definitions of the functions $a \dotplus b$, $a \times b$, $a \approx b$, $a \rightarrow b$, $(\mu \mathrm{v}_h)a$ the factors 4 by 8.

We adjoin further free and bound variables of the new kind, namely

$$\mathrm{A}_h = 8h + 5, \quad \mathrm{V}_h = 8h + 6, \quad h = 1, 2, 3, \ldots$$

and new matrix-forms

$$\mathrm{El}(a, b) = 8J_3(6, a, b) + 3$$

$\mathrm{El}(a, b)$ is to be read "*a is an element of b*".

We adjoin further the operations (EV_h) and (AV_h) which bind the variables of the new type:

$$(\mathrm{EV}_h)a = 8J_3(7, h, a) + 3,$$
$$(\mathrm{AV}_h)a = 8J_3(8, h, a) + 3.$$

With the help of these functions we define the numerical forms, matrix forms, numerical expressions, and matrices in essentially the same way as we defined them in Chapter II. We shall not give these definitions and content ourselves with two examples of matrices of the new system which

are not matrices of (S):

$$(EV_1)(Av_1)El(v_1, V_1),$$
$$(AV_1)(EV_2)(Av_1)[El(v_1, V_2) \leftrightarrow El(v_1 + a, V_1)].$$

The first matrix is to be read thus: *There is a set* V_1 *such that for every integer* v_1, v_1 *is an element of* V_1, and the second: *For every set* V_1 *there is a set* V_2 *such that an arbitrary integer* v_1 *is an element of* V_2 *if and only if* $v_1 + a$ *is an element of* V_1.

The axioms of (S) remain axioms of the new system. We add however new axioms in which variables of the new kind occur. These new axioms are

$$(AV_k)S(A_h, V_k, \Phi) \rightarrow \Phi,$$
$$\Phi \rightarrow (EV_k)S(A_h, V_k, \Phi),$$
$$(EV_k)(Av_l)[El(v_l, V_k) \leftrightarrow S(a_m, v_l, \Psi)],$$
$$(Av_1)[El(v_1, A_h) \leftrightarrow El(v_1, A_k)] \rightarrow [\Theta \leftrightarrow S(A_h, A_k, \Theta)]$$

where Φ is a matrix (in the new sense) in which the variable V_k does not occur, Ψ a matrix in the new sense in which neither the variable V_k nor the variable v_l occur and Θ a wholly arbitrary matrix of the new type.

The rules of proof remain unchanged but we add still two rules which allow us to perform with the quantifiers (EV_h) and (AV_h) the same operations which were described in theorem III 3 2 for the quantifiers (Ev_h) and (Av_h) [6].

The definition of formal proofs remains essentially the same as in the system (S).

This brief description of the new system will be sufficient for our purposes. We shall call the new system (S_1) and shall write $\vdash_1 \Phi$ instead of "Φ *is provable in* (S_1)".

It is easy to see that the semantical notions have their counterparts in the system (S_1). In particular, there exists a matrix Ξ of (S_1) in which a_1 is the only free variable such that Ξ is the

[6] $S(A_h, \Omega, Z)$ denotes here the result of substitution of Ω for the free variable A_h throughout Z. This function is defined similarly as the function $S(i, p, \Phi)$ of Chapter II.

counterpart of the class $\mathfrak{Tr}$. To obtain the matrix Ξ we write down with the help of signs allowed in (S_1) the explicit definition of the class $\mathfrak{Tr}$ given in Chapter IV, section 4, pp. 60—63.

The matrix Ξ has the following important property:

Theorem 2. *If Θ is a matrix of the system* (S) *containing exactly k free variables, then*

$$\vdash_1 \Xi(D(\Theta(D_{n_1}, \ldots, D_{n_k}))) \leftrightarrow \Theta(D_{n_1}, \ldots, D_{n_k}).$$

The proof of this theorem, which proceeds by induction on Θ, is too long to be reproduced here. We shall content ourselves with the special case when Θ is the matrix $a_1 \boldsymbol{+} a_2 \approx a_3$.

From the definition of satisfaction it follows that

$$(1) \qquad \Theta(D_{n_1}, D_{n_2}, D_{n_3}) \in \mathfrak{Tr} \equiv n_1 + n_2 = n_3.$$

The counterparts of the left and the right hand side of this equivalence in the system (S_1) are

$$\Xi(D(\Theta(D_{n_1}, D_{n_2}, D_{n_3}))) \text{ and } \Theta(D_{n_1}, D_{n_2}, D_{n_3}).$$

Since the equivalence (1) can be proved by means formalizable in (S_1), we infer that the equivalence built from the counterparts of the both sides of the equivalence (1) is provable in (S_1). This, however, is the content of theorem 2.

Using theorem 2 we can prove

Theorem 3. $\vdash_1 (\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k)$.

Proof. Taking in theorem 2 $\Theta = (\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k)$, we obtain

$$(2) \qquad \vdash_1 \Xi(D((\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k))) \leftrightarrow (\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k).$$

The left hand side of this equivalence is provable in (S_1). Indeed, we have seen in theorem 1 that the sentence

$$(\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k)$$

is an element of $\mathfrak{Tr}$ and the proof of this theorem can be carried out in the system (S_1). In other words there exists a formalized proof (in (S_1)) of the sentence $\Xi(D((\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k)))$.

The left hand side of (2) being provable in (S_1), we obtain by the propositional calculus

$$\vdash_1 (\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k),$$

and the proof of theorem 3 is complete.

Theorem 3 shows that the sentence $(\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k)$, which is undecidable in (S), becomes decidable in a more comprehensive system (S_1). This shows once more that the undecidability of this sentence is caused not by the difficulty of the problem which it represents but by the inefficiency of the rules of proof admitted in (S).

The same method which we used in the proof of theorem 3 leads to a more general result

Theorem 4. *If Φ is a sentence of the system* (S), $\Phi \in \mathfrak{Tr}$, *and the proof that Φ is in $\mathfrak{Tr}$ is formalizable within* (S_1), *then* $\vdash_1 \Phi$.

It follows from our discussion that many sentences undecidable in (S) become decidable in (S_1). One could think therefore that (S_1) is already an adequate formalization of the intuitive arithmetic. This is, however, not so, because there exist sentences undecidable in (S_1) which are intuitively true in the same degree as is the sentence $(\mathrm{Av}_k) \sim \Phi(D_m, \mathrm{v}_k)$.

To obtain a sentence undecidable in (S_1) we can use one of the methods discussed in Chapter VI. If we use, e.g., the method of VI 2 2 we obtain an undecidable sentence of the form

$$(\mathrm{Av}_k) \sim \Phi_1(D_m, \mathrm{v}_k)$$

where Φ_1 is a recursive definition of the relation

$$\lambda gn\ [g \textit{ represents a formal proof of } n \textit{ in } (S_1)].$$

The new undecidable sentence can be decided in a still more comprehensive system (S_2) which we obtain from (S_1) approximatively in the same way as we obtained (S_1) from (S). At the same time we can construct new undecidable sentences which cannot be decided in (S_2) but become decidable in a new system (S_3). This process can go on indefinitely. The problem how long the construction of different systems can be continued and when (and

whether) we shall reach a stage on which all sentences of (S) will become decidable is at present still far from solution. At any rate two facts are clear: the construction of different systems (S_ν) can be pushed beyond the finite values of ν, and a system (S_ν) in which all sentences of (S) would be decidable, if exists at all, must correspond to an enormous transfinite ordinal ν.

3. Sentences provable simultaneously in (S) and (S_1). We have seen in section 2 that there exist sentences of (S) which are unprovable in (S) but are provable in the larger system (S_1). Gödel discovered the curious fact that the passage from (S) to (S_1) has still another effect: many proofs already existent in (S) can be essentially simplified beyond limits available in (S) [7].

We shall prove here a theorem which is closely related to, but slightly weaker than, the quoted theorem of Gödel. To express this theorem we shall introduce the following notations:

Let $\mathfrak{R}_0$ be a relation which holds between integers n and g if and only if g represents a formal proof of n in the system (S). Let $\mathfrak{R}_1$ be a corresponding relation for the system (S_1). Both $\mathfrak{R}_0$ and $\mathfrak{R}_1$ are recursive relations.

If Φ is provable in (S), then the least g such that $\mathfrak{R}_0(\Phi, g)$ will be called the *minimal* S_0-*proof* of Φ. We define similarly the *minimal* S_1-*proof* of Φ for those Φ which are provable in (S_1).

With these definitions our result can be stated as follows:

Theorem 1. *For every recursive function F there exists a sentence Φ of* (S) *such that* $\vdash \Phi$ *and* $\vdash_1 \Phi$ *and such that the minimal* S_0- *and* S_1-*proofs of* Φ *satisfy the inequality* $g_0 > F(g_1)$.

If we take e.g. $F(x) = 10^{10}\, x$, we can say that there exist sentences of (S) which are provable in (S) and in (S_1) but whose proofs in (S_1) are essentially simpler than in (S) because every proof in (S) is at least 10^{10} times greater than a proof of the same sentence in (S_1) [8].

[7] Gödel [10].

[8] Our measure of simplicity of a proof is thus the value of the integer which represents the proof. One could take other scales of comparison and

In order to prove theorem 1 we assume that it is false and derive a contradiction from this assumption. Let us therefore assume that there exists a recursive function F such that if Φ is a sentence provable both in (S) and in (S_1), then the minimal S_0- and S_1-proofs g_0 and g_1 satisfy the inequality $g_0 \leqslant F(g_1)$.

Let $\mathfrak{R}_i^*$ $(i = 1, 0)$ be relations which hold between matrices of (S) and their minimal S_i-proofs:

$$\mathfrak{R}_i^*(n, g) \equiv \mathfrak{R}_i(n, g) \cdot (q)_g \sim \mathfrak{R}_i(n, q).$$

If Δ_i is a recursive definition of $\mathfrak{R}_i$, then we can take matrices

$$\Delta_i^* = \Delta_i \mathbin{\&} (\mathrm{Av}_p)[\mathrm{v}_p \lhd \mathrm{a}_2 \rightarrow \backsim \Delta_i(\mathrm{a}_1, \mathrm{v}_p)]$$

as recursive definitions of $\mathfrak{R}_i^*$ (p is an arbitrary integer such that v_p does not occur in Δ_i).

Matrices Δ_i^* satisfy evidently the formulas

$$(1) \qquad \vdash_i \Delta_i^*(\mathrm{a}_1, \mathrm{a}_2) \mathbin{\&} \Delta_i^*(\mathrm{a}_1, \mathrm{a}_3) \rightarrow \mathrm{a}_2 \approx \mathrm{a}_3 \quad (i = 1, 0).$$

Let φ be a recursive definition of F, and σ a recursive definition of the function $S(x) = S(1, D_x, x)$. Let us further assume that the bound variables v_h and v_k do not occur in Δ_0^*, Δ_1^*, φ, and σ, and put

$$\Phi = \Delta_1^*(\mathrm{a}_1, \mathrm{a}_2) \mathbin{\&} (\mathrm{Ev}_k)[(\mathrm{v}_k \lhd \varphi(\mathrm{a}_2) + 1) \mathbin{\&} \Delta_0^*(\mathrm{a}_1, \mathrm{v}_k)].$$

The matrix Φ is a recursive definition of the binary relation

$$\mathfrak{R} = \lambda ng[\mathfrak{R}_1^*(n, g) \cdot (\exists q)_{F(n)+1} \mathfrak{R}_0^*(n, q)].$$

We maintain that this relation satisfies the condition

$$(2) \qquad n \in \mathfrak{T} \equiv (\exists g)\mathfrak{R}(n, g).$$

Indeed, if $\mathfrak{R}(n, g)$, then there is a q such that $\mathfrak{R}_0^*(n, q)$, and therefore $\mathfrak{R}_0(n, q)$ which implies that $n \in \mathfrak{T}$. Conversely, let us assume that $n \in \mathfrak{T}$. Hence n is provable in (S) and therefore also

consider e.g. a proof as simpler than the other if it contains less matrices. In this case a theorem similar to theorem 1 is still true: it is precisely the theorem of Gödel quoted in the previous footnote.

in (S_1) since every matrix provable in (S) is at the same time provable in (S_1). Denoting by g_i the minimal S_i-proof of n we have $\mathfrak{R}_i^*(n, g_i)$, and according to our fundamental assumption $g_0 \leqslant F(g_1)$. Hence, $(\exists q)_{F(g_1)+1}\mathfrak{R}_0^*(n, q)$, and we obtain $\mathfrak{R}(n, g_1)$.

Equivalence (2) shows that we can apply theorem VI 2 2 (p. 96) to the relation $\mathfrak{R}$ and the matrix Φ. In this way we obtain an integer $m = S(\Psi)$ where

$$\Psi = (\mathrm{Av}_k) \backsim \Phi(\sigma(\mathrm{a}_1), \mathrm{v}_h)$$

such that

$$(3) \qquad (\mathrm{Av}_h) \backsim \Phi(D_m, \mathrm{v}_h) \text{ non} \in \mathfrak{T},$$

$$(4) \qquad (\mathrm{Av}_h) \backsim \Phi(\sigma(D(\Psi)), \mathrm{v}_h) \leftrightarrow (\mathrm{Av}_h) \backsim \Phi(D_m, \mathrm{v}_h) \in \mathfrak{T},$$

$$(5) \qquad \backsim \Phi(D_m, D_q) \in \mathfrak{T} \text{ for } q = 1, 2, 3, \ldots,$$

$$(6) \qquad m = (\mathrm{Av}_h) \backsim \Phi(\sigma(D(\Psi)), \mathrm{v}_h).$$

Formula (5) shows that $(\mathrm{Av}_h) \backsim \Phi(D_m, \mathrm{v}_h) \in \mathfrak{Tr}$ (cf. Chapter V, section 5, theorem 12, p. 72). According to (6) and (4) we obtain therefore $m \in \mathfrak{Tr}$, and on inspecting the proof of this formula we find that it is formalizable within (S_1). From theorem 2 4 of the Appendix (p. 111) we infer therefore that $\vdash_1 m$, i.e. that there is an integer g_1 such that $\mathfrak{R}_1^*(m, g_1)$. Since Δ_1^* is a recursive definition of $\mathfrak{R}_1^*$ we obtain

$$(7) \qquad \vdash \Delta_1^*(D_m, D_{g_1}).$$

Formulas (3) and (6) prove further that m non $\in \mathfrak{T}$ and hence that $\sim \mathfrak{R}_0^*(m, q)$ for $q = 1, 2, 3, \ldots, F(g_1)$. This gives

$$\vdash \backsim \Delta_0^*(D_m, D_1) \;\&\; \backsim \Delta_0^*(D_m, D_2) \;\&\; \ldots \;\&\; \backsim \Delta_0^*(D_m, D_{F(g_1)}).$$

According to theorem III 7 7 (p. 52) this formula can be written thus:

$$(8) \qquad \vdash (\mathrm{Av}_h)[\mathrm{v}_h \lessdot \varphi(D_{g_1}) + 1 \rightarrow \backsim \Delta_0^*(D_m, \mathrm{v}_h)]$$

since $\vdash \varphi(D_{g_1}) \approx D_{F(g_1)}$ according to the definition of the numerical expression φ.

Observe now that

$$\vdash (\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k) \leftrightarrow (\mathrm{Av}_k)\{\Delta_1^*(D_m, \mathrm{v}_k) \rightarrow$$
$$(\mathrm{Av}_h)[\mathrm{v}_h \lessdot \varphi(\mathrm{v}_k) + 1 \rightarrow \backsim \Delta_0^*(D_m, \mathrm{v}_h)]\}.$$

Using formulas (1), (7), and theorem III 8 1 (p. 54) we obtain from this equivalence

$$\vdash (\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k) \leftrightarrow (\mathrm{Av}_h)[\mathrm{v}_h \lhd \varphi(D_{g_1}) + 1 \rightarrow \backsim \Delta_0^*(D_m, \mathrm{v}_h)]$$

and hence by (8) $\vdash (\mathrm{Av}_k) \backsim \Phi(D_m, \mathrm{v}_k)$. This conclusion contradicts (3). Hence our assumption must be false and theorem 1 is demonstrated.

BIBLIOGRAPHY

[1] R. CARNAP, *Logical syntax of language.* London–New York 1937.

[2] A. CHURCH, *An unsolvable problem of elementary number theory.* American Journal of Mathematics, vol. 58 (1936), pp. 345–363.

[3] ———, *Introduction to mathematical logic.* Annals of mathematics studies, No 13, Princeton 1944.

[4] R. DEDEKIND, *Was sind und was sollen die Zahlen.* Braunschweig 1888.

[5] L. DICKSON, *Introduction to the theory of numbers.* 4-th edition, Chicago 1936.

[6] A. FRAENKEL, *Einleitung in die Mengenlehre.* 2nd edition, Berlin 1928.

[7] G. FREGE, *Grundlagen der Arithmetik.* Breslau 1884.

[8] G. GENTZEN, *Die Widerspruchsfreiheit der reinen Zahlentheorie.* Mathematische Annalen vol. 112 (1936), pp. 493–565.

[9] K. GÖDEL, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I.* Monatshefte für Mathematik und Physik vol. 38 (1931), pp. 173–198.

[10] ———, *Über die Länge der Beweise.* Ergebnisse eines mathematischen Kolloquiums, vol. 7 (1936), pp. 23–24.

[11] D. HILBERT–P. BERNAYS, *Grundlagen der Mathematik.* vol. 1, Berlin 1934, vol. 2, Berlin 1939.

[12] S. C. KLEENE, *General recursive functions of natural numbers.* Mathematische Annalen, vol. 112 (1935–6), pp. 727–742.

[13] ———, *Recursive predicates and quantifiers.* Transactions of the American Mathematical Society, vol. 53 (1943), pp. 41–73.

[14] E. LANDAU, *Grundlagen der Analysis.* Leipzig 1930.

[15] I. L. NOVAK, *A construction of models for consistent systems.* Fundamenta Mathematicae, vol. 37 (1950), pp. 87–110.

[16] E. L. POST, *Recursively enumerable sets of positive integers and their decision problems.* Bulletin of the American Mathematical Society, vol. 50 (1944), pp. 284–316.

[17] W. V. QUINE, *Concatenation as a basis for arithmetic.* Journal of Symbolic Logic, vol. 11 (1946), pp. 105–114.

[18] J. RICHARD, *Les principes des mathématiques et le problème des ensembles.* Revue générale des sciences pures et appliquées, vol. 16 (1905), pp. 541–543.

[19] J. B. ROSSER, *Extensions of some theorems of Gödel and Church.* Journal of Symbolic Logic, vol. 1 (1936), pp. 87–91.

[20] A. TARSKI, *Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften* I. Monatshefte für Mathematik und Physik, vol. 37 (1930), pp. 361–404.

[21] ———, *Sur les ensembles définissables de nombres réels.* Fundamenta Mathematicae, vol. 17 (1930), pp. 210–239.

[22] ———, *Der Wahrheitsbegriff in formalisierten Sprachen.* Studia Philosophica, vol. 1 (1936), pp. 261–405.

[23] ———, *On undecidable statements in enlarged systems of logic and the concept of truth.* Journal of Symbolic Logic, vol. 4 (1939), pp. 105–112.

[24] A. M. TURING, *On computable numbers, with an application to the Entscheidungsproblem.* Proceedings of the London Mathematical Society, vol. 42 (1937), pp. 230–265.

[25] ———, *Systems of logic based on ordinals.* Proceedings of the London Mathematical Society, vol. 45 (1939), pp. 161–228.